East Tennessee State University

Department of Physics, Astronomy & Geology

Astrophysics ASTR3415

Part 3: Applications of General Relativity

These notes discuss some detailed applications of the theory of general relativity developed in Part 2: to the Schwarzschild metric and classical tests of GR; to gravitational waves and to black holes. Lastly, we discuss the Robertson-Walker metric of cosmology, as a precursor to the review of cosmological models in Part 4.

Martin Hendry

October 2005.

Chapter 1

Static Models with Spherical Symmetry

1.1 Orthogonal metrics

In almost all astrophysical situations that we will consider in GR-II, we can generally work with the metric tensor in **orthogonal** form. This means that, in a particular coordinate system, the components, $g_{\alpha\beta}$, of the metric tensor satisfy

$$g_{\alpha\beta} = 0 \quad \text{for all } \alpha \neq \beta \tag{1.1}$$

This implies that there are no 'cross terms' in the expression for the invariant interval; i.e. we can write

$$ds^2 = g_{\alpha\alpha}(dx^\alpha)^2 \tag{1.2}$$

Note that the components of a metric will not be orthogonal in *any* coordinate system. We can prove this as follows.

Suppose that $g_{\alpha\beta}$ are the metric components in a particular coordinate system, such that eq.

(1.1) holds. Let $g'_{\mu\nu}$ denote the metric components in another coordinate system. Since $\mathbf{g}$ is a $(0,2)$ tensor, we have

$$g'_{\mu\nu} = \frac{\partial x^\alpha}{\partial x'^\mu}\frac{\partial x^\beta}{\partial x'^\nu}g_{\alpha\beta} \tag{1.3}$$

Eq. (1.1) implies that

$$g'_{\mu\nu} = \frac{\partial x^\alpha}{\partial x'^\mu}\frac{\partial x^\alpha}{\partial x'^\nu}g_{\alpha\alpha} \tag{1.4}$$

but it does **not** follow that

$$g'_{\mu\nu} = 0 \quad \text{for all } \mu' \neq \nu' \tag{1.5}$$

In fact the question of whether the metric components are orthogonal in a particular coordinate system is closely related to the question of whether the basis vectors are orthogonal in that coordinate system. (Although we often choose a coordinate basis with this property, we do not *have* to make such a choice). Consider, for example, a coordinate system with basis vectors $\{\vec{e}_i\}$, and two vectors $\vec{A} = A^i\vec{e}_i$ and $\vec{B} = B^i\vec{e}_i$.

The scalar product $\vec{A}\cdot\vec{B}$ was defined in GR-I as

$$\vec{A}\cdot\vec{B} = (A^i\vec{e}_i)\cdot(B^j\vec{e}_j) = A^iB^j(\vec{e}_i\cdot\vec{e}_j) \tag{1.6}$$

from which it follows that

$$g_{ij} = \vec{e}_i\cdot\vec{e}_j \tag{1.7}$$

Thus, in particular, $g_{ij} = 0$ if and only if $\vec{e}_i$ and $\vec{e}_j$ are orthogonal.

1.1.1 Contravariant components for an orthogonal metric

It is generally in our interests to *choose* a coordinate system in which the metric coefficients are orthogonal in form. This is because it simplifies the expressions for certain tensors and other geometrical objects in which we are interested.

For example, the contravariant metric components are also orthogonal, and the diagonal terms are simply given by the reciprocal of the covariant diagonal terms. We can prove this as follows. We know that

$$g^{\alpha\beta} g_{\alpha\gamma} = \delta^{\beta}_{\gamma} \tag{1.8}$$

But from eq. (1.1) it follows that

$$g^{\gamma\beta} g_{\gamma\gamma} = \delta^{\beta}_{\gamma} \tag{1.9}$$

i.e.

$$g^{\gamma\beta} = 0 \quad \text{if} \gamma \neq \beta \tag{1.10}$$

and

$$g^{\gamma\gamma} = 1/g_{\gamma\gamma} \tag{1.11}$$

1.1.2 Christoffel symbols for an orthogonal metric

One can show that the Christoffel symbols also take a simple form for an orthogonal metric:-

$$\begin{aligned}
\Gamma^{\lambda}_{\mu\nu} &= 0 \quad \text{for } \lambda,\ \mu,\ \nu \text{ all different} \\
\Gamma^{\lambda}_{\lambda\mu} &= \Gamma^{\lambda}_{\mu\lambda} = g_{\lambda\lambda,\mu}/2g_{\lambda\lambda} \\
\Gamma^{\lambda}_{\mu\mu} &= -g_{\mu\mu,\lambda}/2g_{\lambda\lambda} \\
\Gamma^{\lambda}_{\lambda\lambda} &= g_{\lambda\lambda,\lambda}/2g_{\lambda\lambda}
\end{aligned} \tag{1.12}$$

(Note: the summation convention does not apply in these equations). The proof of these results is left as an exercise; see Homework Assignment 2.

1.1.3 Geodesic equations for an orthogonal metric

We can find another (and potentially more useful) form of the geodesic equation. For affine parameter, p, this takes the form

$$\frac{d}{dp}\left(g_{\lambda\nu}\frac{dx^\nu}{dp}\right) - \frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x^\lambda}\frac{dx^\mu}{dp}\frac{dx^\nu}{dp} = 0 \tag{1.13}$$

For an orthogonal metric this reduces further to

$$\frac{d}{dp}\left(g_{\lambda\lambda}\frac{dx^\lambda}{dp}\right) - \frac{1}{2}\frac{\partial g_{\mu\mu}}{\partial x^\lambda}\left(\frac{dx^\mu}{dp}\right)^2 = 0 \tag{1.14}$$

The proof of equations (1.13) and (1.14) is left as an exercise.

1.2 Geometrised units

In GR-I we adopted units in which the speed of light, $c = 1$. This effectively means that we are measuring time in units of **length** – specifically, the distance travelled by light in that time. Thus

$$1 \text{ second} \equiv 3 \times 10^8 \text{ m}$$

At the end of GR-I we derived the weak field limit of Einstein's equations, which allowed us to determine the constant, k, in equation (1.1) in terms of the gravitational constant, G. Now, in SI units

$$G \simeq 6.67 \times 10^{-11}\,\mathrm{N\,m^2\,kg^{-2}}$$

but the Newton is a composite SI unit; in fact

$$1\,\mathrm{N} = 1\,\mathrm{kg\,m\,s^{-2}}$$

so that

$$G \simeq 6.67 \times 10^{-11}\,\mathrm{m^3\,kg^{-1}\,s^{-2}}$$

Replacing our unit of time with the unit of length defined above, we obtain

$$G \simeq 7.41 \times 10^{-28}\,\mathrm{m}\,\mathrm{kg}^{-1}$$

In GR-II we will, in turn, find it useful to set $G = 1$; this effectively means that we will measure *mass* also in units of length. This approach is often referred to as **geometrised units**. It follows that, in these new units

$$1\ \mathrm{kg} = 7.41 \times 10^{-28}\ \mathrm{m}$$

In summary, our geometrised units take the form

Unit of length:	1 m		
Unit of time:	1 m	$\equiv$	3.33×10^{-9} s
Unit of mass:	1 m	$\equiv$	1.34×10^{27} kg

1.3 Spherically symmetric metrics

A large part of GR-I was concerned with formulating Einstein's equations – having first to develop a lot of necessary mathematical machinery to make this possible. In GR-II we now face the (perhaps more interesting) challenge of trying to solve Einstein's equations. This is far from trivial, however, and indeed exact solutions have been obtained for a (fairly small) number of special cases only. Fortunately however, these cases include some examples which are excellent approximations to real astrophysical situations.

Here we consider the case of a **spherically symmetric solution**, which we will then use to model the spacetime outside and inside a star. What exactly do we mean by spherical symmetry in curved spacetime, however? In the flat Minkowski spacetime of SR we can change our spatial variables to the usual polar coordinates, to give an invariant interval

$$ds^2 = -dt^2 + dr^2 + r^2 \left(d\theta^2 + \sin^2\theta\, d\phi^2\right) \tag{1.15}$$

Thus, surfaces of constant r and t have the geometry of a 2-sphere, with interval

$$d\ell^2 = r^2 \left(d\theta^2 + \sin^2\theta \, d\phi^2 \right) \tag{1.16}$$

We can also use this property to define spherical symmetry in a curved spacetime:

A spacetime is spherically symmetric if every point in the spacetime lies on a 2-D surface which is a 2-sphere

If we label the coordinates of our spacetime by (r', t, θ, ϕ) – the reasons for our choice of labels will become clear shortly – then every point in a spherically symmetric spacetime lies on a 2-D surface which is a 2-sphere, with interval given by

$$d\ell^2 = f(r', t) \left[d\theta^2 + \sin^2\theta \, d\phi^2 \right] \tag{1.17}$$

where $\sqrt{f(r', t)}$ is the **radius of curvature** of the 2-sphere.

Unlike flat spacetime, where transforming to ordinary polar coordinates immediately reveals the radius of curvature to be simply our radial polar coordinate, r, in curved spacetime no such trivial relation always exists between the angular coordinates of the 2-D sphere and the remaining two coordinates at each point in spacetime.

We can, however, simply *define* a new radial coordinate, r, which satisfies

$$r^2 = f(r', t) \tag{1.18}$$

and we can 'line up' the origins of our 2-sphere coordinate systems, (θ, ϕ), for points in spacetime with different values of r.

Spherical symmetry also requires that any radial path in the space is **orthogonal** to the 2-D spheres on the which the points along that radial path lie, since otherwise this would allow us to define a preferred direction in the space. This implies that, in the metric of our spherically symmetric spacetime

$$g_{r\theta} = g_{r\phi} = 0 \tag{1.19}$$

Hence, we have restricted the form of the spacetime metric to

$$ds^2 = g_{tt}dt^2 + 2g_{tr}drdt + 2g_{t\theta}d\theta dt + 2g_{t\phi}dtd\phi + g_{rr}dr^2 + r^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right) \tag{1.20}$$

Consider now the curve $r =$ const., $\theta =$ const., $\phi =$ const. This is the wordline of a particle in the spacetime which has constant spatial coordinates. This curve must *also* be orthogonal to the 2-spheres on which each point on the curve lies, since otherwise we could define a preferred direction in *spacetime*. This means that we can further restrict our metric coefficients to be

$$g_{t\theta} = g_{t\phi} = 0 \tag{1.21}$$

So we can write the general form of the metric for a spherically symmetric spacetime as

$$ds^2 = g_{tt}dt^2 + 2g_{tr}drdt + g_{rr}dr^2 + r^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right) \tag{1.22}$$

where g_{tt}, g_{tr} and g_{rr} are arbitrary functions of r and t.

1.3.1 The case of a static spacetime

We now consider the special case of a **static spherically symmetric spacetime** (which we henceforth refer to as S^4 for short). This means that we can find a time coordinate, t, for which

1. all metric components are independent of t
2. the metric is unchanged if we apply the transformation $t \to -t$ (i.e. if we undergo time reversal)

It is fairly easy to see that property (2) implies that $g_{tr} = 0$, which means that we can write the interval for S^4 as

$$ds^2 = -e^{\nu}dt^2 + e^{\lambda}dr^2 + r^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right) \tag{1.23}$$

which is orthogonal in form. Note that we have introduced the functions $\nu(r)$ and $\lambda(r)$ in replacing g_{tt} and g_{rr}. Since the exponential function is strictly positive for all r, this replacement is legitimate provided that $g_{tt} < 0$ and $g_{rr} > 0$ for all points in our spacetime. We will consider carefully later whether these conditions are satisfied, and under what circumstances they break down.

1.3.2 Christoffel symbols for S^4

Using equations (1.12) it is straightforward to calculate the Christoffel symbols for this metric. Denoting by dashes differentiation with respect to r, the Christoffel symbols are given by

$$\begin{array}{rcc@{\qquad}rcc}
\Gamma^t_{rt} = \Gamma^t_{tr} & = & \frac{1}{2}\nu' & \Gamma^\theta_{r\theta} = \Gamma^\theta_{\theta r} & = & \frac{1}{r} \\
\Gamma^r_{tt} & = & \frac{1}{2}\nu' e^{\nu-\lambda} & \Gamma^\theta_{\phi\phi} & = & -\sin\theta\cos\theta \\
\Gamma^r_{rr} & = & \frac{1}{2}\lambda' & \Gamma^\phi_{r\phi} = \Gamma^\phi_{\phi r} & = & \frac{1}{r} \\
\Gamma^r_{\theta\theta} & = & -re^{-\lambda} & \Gamma^\phi_{\theta\phi} = \Gamma^\phi_{\phi\theta} & = & \cot\theta \\
\Gamma^r_{\phi\phi} & = & -re^{-\lambda}\sin^2\theta & & &
\end{array}$$

$$\text{All others zero} \tag{1.24}$$

1.3.3 Ricci tensor for S^4

We can write the Ricci tensor as

$$R_{\lambda\nu} = \Gamma^\tau_{\lambda\nu}\Gamma^\sigma_{\tau\sigma} - \Gamma^\tau_{\lambda\sigma}\Gamma^\sigma_{\tau\nu} + \Gamma^\sigma_{\lambda\nu,\sigma} - \Gamma^\sigma_{\lambda\sigma,\nu} \tag{1.25}$$

Substituting the results of equations (1.24) into equation (1.25) we find that

$$R_{tt} = \frac{1}{2}e^{\nu-\lambda}\left(\nu'' + \frac{1}{2}\nu'^2 - \frac{1}{2}\nu'\lambda' + \frac{2}{r}\nu'\right) \tag{1.26}$$

$$R_{rr} = -\frac{1}{2}\left(\nu'' + \frac{1}{2}\nu'^2 - \frac{1}{2}\nu'\lambda' - \frac{2}{r}\lambda'\right) \tag{1.27}$$

$$R_{\theta\theta} = 1 - e^{-\lambda}\left[1 + \frac{r}{2}\left(\nu' - \lambda'\right)\right] \tag{1.28}$$

$$R_{\phi\phi} = R_{\theta\theta}\sin^2\theta \tag{1.29}$$

and all other terms of the Ricci tensor are identically zero.

1.4 Derivation of the Schwarzschild metric

We now seek to apply the general S^4 metric to derive the spacetime exterior to a spherically symmetric star; this is known as the **Schwarzschild solution**. If the star is in an isolated region of space, then we can assume that all components of the Ricci tensor exterior to the star are identically zero. Hence the right hand side of equations (1.26) – (1.29) are all zero. Thus

$$e^{\lambda-\nu}R_{tt} + R_{rr} = \frac{\nu' + \lambda'}{r} = 0 \tag{1.30}$$

which in turn implies that

$$\nu + \lambda = \text{constant} \tag{1.31}$$

At large distances from the star we want the Schwarzschild metric to reduce to SR. Hence, as

$$r \to \infty\ , \quad e^{\nu} \to 1 \quad \text{and} \quad e^{\lambda} \to 1 \tag{1.32}$$

Thus, as

$$r \to \infty\ , \quad \nu \to 0 \quad \text{and} \quad \lambda \to 0 \tag{1.33}$$

which implies that

$$\nu + \lambda = 0 \tag{1.34}$$

so that

$$e^{\nu} = e^{-\lambda} \tag{1.35}$$

This allows us to eliminate ν from equation (1.28), giving

$$e^{-\lambda}\left(1 - \lambda' r\right) = 1 \tag{1.36}$$

i.e.

$$\frac{d}{dr}\left(re^{-\lambda}\right) = 1 \tag{1.37}$$

which we can integrate to give

$$e^{\nu} = e^{-\lambda} = 1 + \frac{\alpha}{r} \tag{1.38}$$

where α is a constant.

To evaluate α, suppose we release a material 'test' particle (i.e. a particle of so little rest mass that it does not disturb the spacetime metric) from rest. Thus, initially

$$\frac{dx^j}{d\tau} = 0 \quad \text{for } j = 1, 2, 3 \tag{1.39}$$

where τ is **proper time**, and

$$\frac{dx^0}{d\tau} \equiv \frac{dt}{d\tau} \neq 0 \tag{1.40}$$

After some further reduction we see that

$$\frac{dt}{d\tau} = e^{-\nu/2} \tag{1.41}$$

We now apply the first of the geodesic differential equations. At the instant when the particle is released this reduces to

$$\frac{d^2 r}{d\tau^2} + \Gamma^r_{tt}\left(\frac{dt}{d\tau}\right)^2 = 0 \tag{1.42}$$

Substituting from equations (1.24) and (1.41) we obtain finally

$$\frac{d^2r}{d\tau^2} = \frac{\alpha}{2r^2} \tag{1.43}$$

In the limit of a weak gravitational field this result must reduce to the prediction of Newtonian gravity, which predicts

$$\frac{d^2r}{dt^2} = -\frac{GM}{r^2} \tag{1.44}$$

where M is the mass of the star. If we adopt convenient units such that the gravitational constant, $G = 1$ (see below), this means that

$$\alpha = -2M \tag{1.45}$$

We can now write down the invariant interval for the Schwarzschild metric:

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dt^2 + \frac{dr^2}{\left(1 - \frac{2M}{r}\right)} + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2 \tag{1.46}$$

1.5 The Schwarzschild radius and the event horizon

The quantity $2M$ in equation (1.46) is known as the **Schwarzschild radius**. The Sun (which has a mass of about 2×10^{30} kg) has a Schwarzschild radius of about 3 km.

We can see from equation (1.46) that the metric 'misbehaves' when $r = 2M$, since $g_{tt} = 0$ and $g_{rr} \rightarrow \infty$. We refer to this surface as the **event horizon** of the Schwarzschild metric, and we will discuss its physical significance in greater detail later in the course when we consider black holes. For now, we need only note that – provided $r > 2M$ – the metric is perfectly well behaved. For most stars (like the Sun, for example!) the Schwarzschild radius is much smaller than the physical radius (i.e. the photosphere) of the star. When this is the case we can certainly apply the Schwarzschild metric in the form of equation (1.46) to describe the spacetime exterior to the surface of the star; we do exactly this in the next chapter.

Chapter 2

The Schwarzschild Metric and Classical Tests of GR

In this chapter we apply the Schwarzschild metric to describe the spacetime exterior to a star. This will allow us to investigate four **Classical Tests of General Relativity**. These are:-

1. The advance of pericentre of planetary orbits
2. Gravitational light deflection
3. Gravitational redshift
4. Gravitational time delay

Before we consider each of these tests in turn, we first determine the equations of geodesics for the Schwarzschild metric.

2.1 Geodesics for the Schwarzschild metric

The geodesics for a material 'test' particle in the Schwarzschild metric satisfy equation (1.13), with the proper time, τ, as affine parameter:

$$\frac{d}{dp}\left(g_{\lambda\nu}\frac{dx^\nu}{dp}\right) - \frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x^\lambda}\frac{dx^\mu}{dp}\frac{dx^\nu}{dp} = 0 \tag{2.1}$$

Notice that the metric coefficients of equation (1.46) are independent of both t and ϕ; hence if we set $\lambda = 0$ and $\lambda = 3$ then the second term on the left hand side of equation (2.1) vanishes. Given also that the Schwarzschild metric is *orthogonal*, it follows that

$$\frac{d}{d\tau}\left(g_{tt}\frac{dt}{d\tau}\right) = \frac{d}{d\tau}\left(g_{\phi\phi}\frac{d\phi}{d\tau}\right) = 0 \tag{2.2}$$

Integrating this gives us

$$g_{tt}\frac{dt}{d\tau} = \text{ constant} \tag{2.3}$$

and

$$g_{\phi\phi}\frac{d\phi}{d\tau} = \text{ constant} \tag{2.4}$$

The geodesic equation for θ (i.e. $\lambda = 2$) follows from equation (1.14):

$$\frac{d}{d\tau}\left(r^2\frac{d\theta}{d\tau}\right) - \frac{1}{2}\frac{\partial}{\partial\theta}\left(r^2\sin^2\theta\right)\left[\frac{d\phi}{d\tau}\right]^2 = 0 \tag{2.5}$$

which reduces to

$$r^2\frac{d^2\theta}{d\tau^2} + 2r\frac{dr}{d\tau}\frac{d\theta}{d\tau} - r^2\sin\theta\cos\theta\left(\frac{d\phi}{d\tau}\right)^2 = 0 \tag{2.6}$$

Equation (2.6) has a particular solution $\theta = \pi/2$; adopting this solution is equivalent to choosing the plane of the orbit of our material particle (e.g. a planet) to lie in the equatorial plane of our coordinate system. Thus

$$\frac{d\theta}{d\tau} = 0 \tag{2.7}$$

Making use of $\theta = \pi/2$ in simplifying equations (2.3) and (2.4), it follows that

$$\frac{dt}{d\tau} = \frac{k}{1 - \frac{2M}{r}} \tag{2.8}$$

and

$$\frac{d\phi}{d\tau} = \frac{h}{r^2} \tag{2.9}$$

where h and k are constants.

We can now use equations (2.8), (2.9) and (2.7) to obtain the geodesic differential equation for r, not directly from equation (1.14) but from equation (**??**)

$$-1 = g_{tt}\left(\frac{dt}{d\tau}\right)^2 + g_{rr}\left(\frac{dr}{d\tau}\right)^2 + g_{\phi\phi}\left(\frac{d\phi}{d\tau}\right)^2 \tag{2.10}$$

which in turn reduces to

$$\left(\frac{dr}{d\tau}\right)^2 = k^2 - 1 - \frac{h^2}{r^2} + \frac{2M}{r}\left(1 + \frac{h^2}{r^2}\right) \tag{2.11}$$

2.2 Planetary orbits in Newtonian theory

The Newtonian equations of motion for a test mass orbiting a mass M (where the orbital plane is taken to be the equatorial plane $\theta = \pi/2$) take the form

$$r^2\frac{d\phi}{dt} = h \tag{2.12}$$

and

$$\frac{d^2r}{dt^2} = -\frac{M}{r^2} + r\left(\frac{d\phi}{dt}\right)^2 \tag{2.13}$$

Note that equations (2.9) and (2.12) are equivalent, apart from the use of proper time, τ, and coordinate time, t, respectively; in Newtonian dynamics, however, there is no distinction between τ and t.

It is customary to solve equation (2.12) by first changing the dependent variable from r to $u = 1/r$, and the independent variable from t to ϕ. Note that

$$\frac{d\phi}{dt} = hu^2 \tag{2.14}$$

and so

$$\frac{dr}{dt} = \frac{d}{dt}\left(\frac{1}{u}\right) = -\frac{1}{u^2}\frac{du}{d\phi}\frac{d\phi}{dt} = -h\frac{du}{d\phi} \tag{2.15}$$

Re-expressing equation (2.13) in terms of u and ϕ gives

$$\frac{d^2u}{d\phi^2} = -u + \frac{M}{h^2} \tag{2.16}$$

This equation has solution

$$u = \frac{M}{h^2}\,(1 + e\cos\phi) \tag{2.17}$$

which represents an ellipse, with eccentricity e, semi-major axis a and focus at $r = 0$. The constant h is related to the **semi-latus rectum**, ℓ of the ellipse, viz:-

$$\ell = \frac{h^2}{M} = a(1 - e^2) \tag{2.18}$$

2.3 The advance of pericentre in GR

We can manipulate equation (2.11) in a similar manner to the solution of the Newtonian differential equation (2.15), by changing the dependent variable from r to u and the independent variable from τ to ϕ. Substituting into equation (2.11) this gives

$$h^2\left(\frac{du}{d\phi}\right)^2 = \left(k^2 - 1\right) - h^2u^2 + 2Mu\left(1 + h^2u^2\right) \tag{2.19}$$

Differentiating this equation and cancelling the common factor of $\mathrm{d}u/\mathrm{d}\phi$ gives the result

$$\frac{d^2u}{d\phi^2} = -u + \frac{M}{h^2} + 3Mu^2 \tag{2.20}$$

Comparing equations (2.16) and (2.20) we see that the effect of a GR treatment is to add the extra term $3Mu^2$ on the right hand side. We will now determine the impact of this extra term on the orbital path of a planet moving around a star of mass, M. For typical planetary orbits in the Solar System this extra term is tiny compared with the second term on the right hand side of equation (2.20); e.g. for the Earth's orbit the ratio

$$\frac{3Mu^2}{M/h^2} \simeq 3 \times 10^{-8} \tag{2.21}$$

Hence, because the extra GR term is very small anyway, we can obtain a very good approximation to equation (2.20) by replacing u in the u^2 term on the right hand side by the solution to the *Newtonian* version of this equation, as given by equation (2.17). Thus,we obtain

$$\frac{d^2u}{d\phi^2} = -u + \frac{M}{h^2} + 3\frac{M^3}{h^4}\left(1 + 2e\cos\phi + e^2\cos^2\phi\right) \tag{2.22}$$

We can write u as the sum of a 'Newtonian' and 'GR' part, i.e.

$$u = u_{\rm N} + u_{\rm GR} \tag{2.23}$$

so that $u_{\rm GR}$ describes the correction to the Newtonian orbit. Subtracting off the Newtonian solution, equation (2.22) becomes

$$\frac{d^2u_{\rm GR}}{d\phi^2} = -u_{\rm GR} + 3\frac{M^3}{h^4}\left(1 + 2e\cos\phi + e^2\cos^2\phi\right) \tag{2.24}$$

Noting that

$$\cos^2\phi = \frac{1}{2}\left(1 + \cos 2\phi\right) \tag{2.25}$$

we can rewrite equation (2.24) as

$$\frac{d^2u_{\rm GR}}{d\phi^2} + u_{\rm GR} = 3\frac{M^3}{h^4}\left(1 + \frac{e^2}{2} + 2e\cos\phi + \frac{e^2}{2}\cos 2\phi\right) \tag{2.26}$$

The right hand side of equation (2.26) takes the form

$$A + B\cos\phi + C\cos 2\phi \tag{2.27}$$

where A, B and C are constants. It is easy to verify that particular integrals for each of these terms are, respectively

$$u_{\mathrm{GR}} = A \tag{2.28}$$

$$u_{\mathrm{GR}} = \frac{1}{2} B\phi \sin\phi \tag{2.29}$$

$$u_{\mathrm{GR}} = -\frac{1}{3} C \cos 2\phi \tag{2.30}$$

and the correction to the Newtonian orbit is given by the sum of these three particular integrals. Since each of the constants, A, B and C is of order the tiny constant M^3/h^4, we see that the first and third terms – given by equations (2.28) and (2.30) – add to the Newtonian solution respectively a completely negligible constant and an equally negligible constant plus a tiny "wiggle".

The second term, on the other hand, is of a different form. Although the constant, B, is negligibly small, the presence of the ϕ in equation (2.29) means that this term produces a continually increasing – and thus ultimately non-negligible – effect. From equations (2.17), (2.23), (2.26) and (2.29) we can obtain

$$u = \frac{M}{h^2}\left(1 + e\cos\phi + \frac{3M^2}{h^2} e\phi \sin\phi\right) \tag{2.31}$$

Now given that $3M^2/h^2$ is very small, and then using the approximations $\cos\beta \simeq 1$ and $\sin\beta \simeq \beta$ for small angle β, and the addition formula

$$\cos(\alpha - \beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta \tag{2.32}$$

we can re-cast equation (2.31) as

$$u = \frac{M}{h^2}\left[1 + e\cos\left(1 - \frac{3M^2}{h^2}\right)\phi\right] \tag{2.33}$$

Comparing equation (2.33) with its Newtonian analogue, equation (2.17), we see that again the solution is elliptical in form and that u (and hence r) is a periodic function of ϕ. Notice, however, that the period, P, is given by

$$P = \frac{2\pi}{1 - 3M^2/h^2} > 2\pi \tag{2.34}$$

This means that the values of r trace out an approximate ellipse, but do not begin to repeat until *after* the radius vector has made a complete revolution. In other words the orbit can be regarded as an ellipse that 'precesses' – as shown in Figure 1, so that the pericentre line advances each orbit by an amount, Δ, given by

$$\Delta = 2\pi \left(1 - \frac{3M^2}{h^2}\right)^{-1} - 2\pi \simeq \frac{6\pi M^2}{h^2} = \frac{6\pi M}{a(1-e^2)} \tag{2.35}$$

If we apply equation (2.35) to the orbit of Mercury, we obtain a perihelion advance which builds up to about 43 seconds of arc per century.

2.3.1 GR's first major success

It had been realised since the mid-19th century – when the existence of the planet Neptune was predicted by Adams and Le Verrier from studying its perturbing effect on the orbit of Uranus – that there was something wrong with the Newtonian predictions for the orbit of Mercury. Le Verrier, applying Newtonian perturbation theory, predicted the existence of another planet inside that of Mercury and even gave it a name – Vulcan. He calculated the orbit which Vulcan was required to have, in order to explain the discrepancy between Mercury's observed orbit and the predictions of Newtonian gravity – after accounting for the perturbing effects of all of the known planets. All attempts to observe Vulcan, however, met with failure.

Einstein's publication of General Relativity in 1916 provided the answer to the mystery of Mercury's orbit. The GR prediction of a perihelion advance matched extremely well the observed discrepancy in Mercury's orbit, and represented the first major success of the theory over Newtonian gravity.

2.3.2 The Binary Pulsar

A much clearer example of pericentre advance can be seen in the **binary pulsar** system PSR 1913+16, discovered in 1974. The orbital period of the system is $P = 0.323$ days, and the periastron is advancing at the rate of more than 4° per year. It is very likely that both members of the binary system are neutron stars. This system has proved to be an excellent laboratory for testing predictions of general relativity and we will consider it again later in this chapter, and in discussing gravitational radiation.

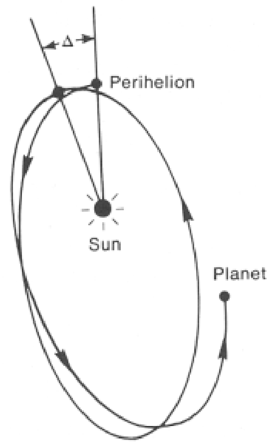


Figure 1

Advance of perihelion, as predicted by General Relativity, resulting in a planetary orbit which is not a closed ellipse, but which precesses slowly.

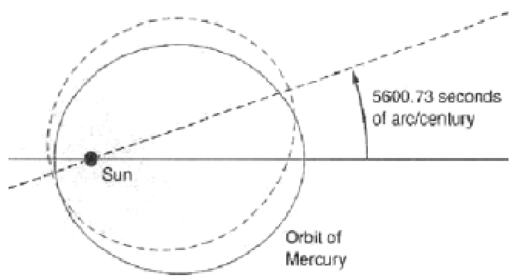


Figure 2

Advance of perihelion of the planet Mercury. All but 43 arcseconds per century of the observed advance of perihelion could be explained simply by Newtonian perturbations by the other known planets; GR provided an explanation for the remaining 43 arcseconds discrepancy. This additional advance is due to the effect of the curvature of spacetime in the vicinity of the Sun: loosely, one can think of this curvature as an additional source of mass-energy which adds to the perturbation on the planet.

2.4 Gravitational light deflection

The second classical test of General Relativity was the deflection of a light ray passing close to a massive object. In GR-I we showed, by considering a lift in free fall, that light deflection is implied by the Strong Principle of Equivalence. In this chapter we examine the effect more formally, within the framework of the Schwarzschild metric.

This classical test attracted much attention in the years immediately following the publication of General Relativity because it was able to be carried out experimentally, by measuring the apparent shift in angular position of stars very close to the limb of the Sun during a total solar eclipse. In 1919 the British astronomer Arthur Eddington led an expedition to the Southern Hemisphere to carry out this test.

It is sometimes written than General Relativity predicts gravitational light deflection while Newtonian gravitation does not. This is only partly true, however. If we regard photons (as, indeed, modern physics holds true) as particles with zero rest mass, then formally they are 'immune' to Newton's gravitational force. If, on the other hand, we regard photons as having a negligible but non-zero mass then – even within a purely Newtonian framework – we can calculate the predicted deflection angle as light passes close to a massive object. In fact, this calculation was first carried out in 1801 by Söldner. Before considering the General Relativistic calculation, it is instructive to derive this Newtonian deflection angle result.

2.4.1 Newtonian light deflection

We consider the path of a photon passing close to a mass, M. In Newtonian dynamics the orbit of the photon during the encounter is a hyperbola, with M at one focus, given by

$$r(\phi) = \frac{r_{\min}(e+1)}{1 + e\cos\phi} \tag{2.36}$$

where $e > 1$ is the eccentricity and $r_{\min}$ is the distance of the photon from M at its point of closest approach. This trajectory is shown in Figure 3. The asymptotic directions of the photon before and after the deflecting encounter are given by

$$\phi = \pm\left(\frac{\pi}{2} + \frac{\Delta\phi}{2}\right) \tag{2.37}$$

where $\Delta\phi$ is the total deflection angle (compared with the undeflected trajectory, from $\phi = -\pi/2$ to $\phi = \pi/2$).

As for a planetary elliptical orbit, the motion of the Newtonian photon satisfies

$$r^2 \frac{d\phi}{dt} = h \tag{2.38}$$

where h is a constant, related to the semi-latus rectum, ℓ, of the hyperbola via:

$$h^2 \; = \; M\ell \; = \; Ma(e^2 - 1) \; = \; M\, r_{\text{min}}(e+1) \tag{2.39}$$

The photon also satisfies an **energy equation**

$$\frac{1}{2}v^2 - \frac{M}{r} = \frac{1}{2}\left[\left(\frac{dr}{dt}\right)^2 + r^2\left(\frac{d\phi}{dt}\right)^2\right] - \frac{M}{r} = E_{\text{tot}} \tag{2.40}$$

where the constant E_{tot} is the total energy, equal to the sum of the kinetic and potential energy (remembering that $G = 1$), per unit mass. In equation (2.40) the two terms in the square brackets are, respectively, the squared radial and transverse velocity components.

Writing

$$\frac{dr}{dt} = \frac{dr}{d\phi}\frac{d\phi}{dt} \tag{2.41}$$

and substituting from equation (2.38) it follows that

$$\frac{h^2}{2r^2}\left[\frac{(dr/d\phi)^2}{r^2} + 1\right] - \frac{M}{r} = E_{\text{tot}} \tag{2.42}$$

From equation (2.36)

$$\frac{dr}{d\phi} = \frac{r_{\text{min}}(e+1)e\sin\phi}{(1+e\cos\phi)^2} = \frac{r^2 e\sin\phi}{r_{\text{min}}(e+1)} \tag{2.43}$$

Substituting into equation (2.42), and after a little algebra we obtain

$$E_{\text{tot}} = \frac{M(e-1)}{2r_{\text{min}}} \tag{2.44}$$

From equation (2.36) we see that $r \to \infty$ when $\cos\phi = -1/e$. Also, setting $v = c = 1$ for $r \to \infty$, from equation (2.40) we see that

$$E_{\text{tot}} = \frac{1}{2} \tag{2.45}$$

Rearranging equation (2.44) it follows that, since $e >> 1$

$$e = 1 + \frac{r_{\rm min}}{M} \simeq \frac{r_{\rm min}}{M} \tag{2.46}$$

Hence the asymptotic direction of the outgoing photon is

$$\cos\phi = \cos\left(\frac{\pi}{2} + \frac{\Delta\phi}{2}\right) = -\sin\left(\frac{\Delta\phi}{2}\right) = -\frac{M}{r_{\rm min}} \tag{2.47}$$

Since $\Delta\phi << 1$,

$$\Delta\phi = \frac{2M}{r_{\rm min}} \tag{2.48}$$

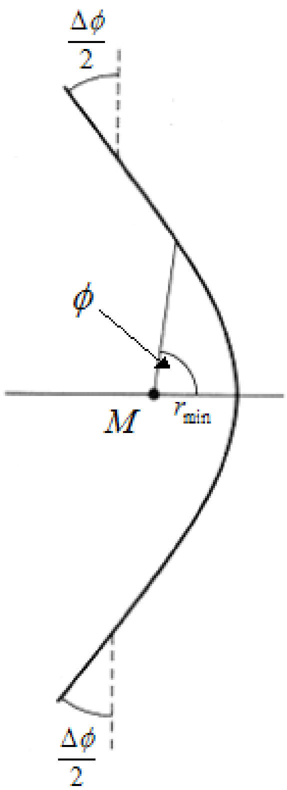


Figure 3

Trajectory of a deflected light ray passing close to a mass, M

The total deflection angle is seen to be twice the angle between the asymptote of the trajectory and the dashed line (the undeflected trajectory)

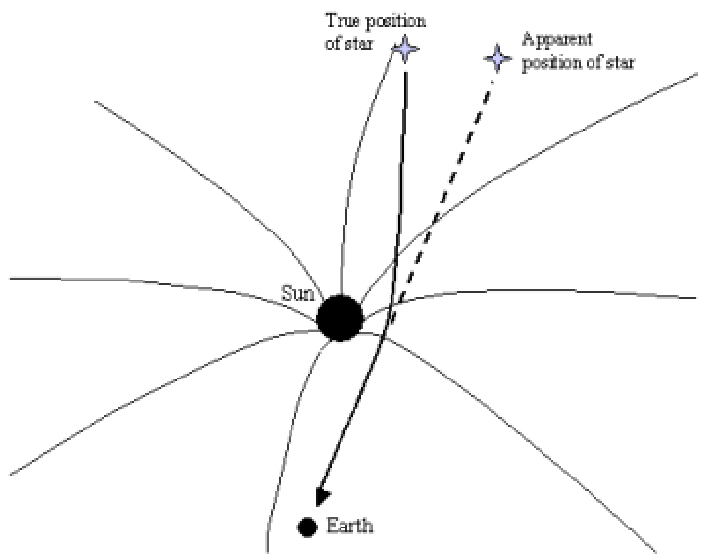


Figure 4

Schematic diagram showing the deflection of a light ray from a background star as it passes close to the Sun. This gravitational light deflection can be observed during a total solar eclipse, and is also seen in radio observations of distant quasars when they lie close to the Sun on the sky.

2.4.2 Light deflection in General Relativity

The geodesics for a photon in the Schwarzschild metric may be derived in a similar manner to Section 2.1, but we now must introduce a new affine parameter, λ (say), since the proper time for a photon is zero.[1]

For the 't' and 'ϕ' geodesic equations it is straightforward to see that we again obtain equations of the form

$$\frac{dt}{d\lambda} = \frac{k}{1 - 2M/r} \tag{2.49}$$

$$\frac{d\phi}{d\lambda} = \frac{h}{r^2} \tag{2.50}$$

since, for the 'θ' equation, we can again spot the particular solution $\theta = \pi/2$. It obviously then also follows that

$$\frac{d\theta}{d\lambda} = 0 \tag{2.51}$$

We can then obtain

$$\left(\frac{dr}{d\lambda}\right)^2 = k^2 - \frac{h^2}{r^2} + \frac{2Mh^2}{r^3} \tag{2.52}$$

We now proceed as in Section 2.3, replacing the dependent variable, r by $u = 1/r$, and the independent variable λ by ϕ. This gives us

$$\frac{d^2u}{d\phi^2} + u = 3Mu^2 \tag{2.53}$$

If we ignore the term on the right hand side we can see that a particular integral is

$$u = \frac{\cos\phi}{r_{\rm min}} \tag{2.54}$$

Following the same approach as Section 2.3, we can obtain a very good approximation to equation (2.53) by replacing u on the right hand side by equation (2.54). This gives the

[1] In fact, the choice of affine parameter will not be important, since we will determine the trajectory of the photon with the coordinate ϕ as the independent variable

equation

$$\frac{d^2u}{d\phi^2} + u = \frac{3M}{r_{\text{min}}^2}\cos^2\phi = \frac{3M}{2r_{\text{min}}^2}\left(1 + \cos 2\phi\right) \tag{2.55}$$

It is straightforward to verify that a particular integral of this approximation is

$$u = \frac{3M}{2r_{\text{min}}^2}\left(1 - \frac{1}{3}\cos 2\phi\right) \tag{2.56}$$

from which it follows that the general solution of equation (2.53) is

$$u = \frac{\cos\phi}{r_{\text{min}}} + \frac{3M}{2r_{\text{min}}^2}\left(1 - \frac{1}{3}\cos 2\phi\right) \tag{2.57}$$

Using equation (2.37) we can rewrite this, for e.g. the outgoing photon trajectory, as

$$u = \frac{\cos\left(\frac{\pi}{2} + \frac{\Delta\phi}{2}\right)}{r_{\text{min}}} + \frac{3M}{2r_{\text{min}}^2}\left[1 - \frac{1}{3}\cos\left(\pi + \Delta\phi\right)\right] \tag{2.58}$$

or

$$u = -\frac{\sin(\Delta\phi/2)}{r_{\text{min}}} + \frac{3M}{2r_{\text{min}}^2}\left[1 + \frac{1}{3}\cos\Delta\phi\right] \tag{2.59}$$

which further simplifies, since $\Delta\phi << 1$, to

$$u = -\frac{\Delta\phi}{2r_{\text{min}}} + \frac{2M}{r_{\text{min}}^2} \tag{2.60}$$

Setting $u = 0$ (i.e. $r \to \infty$) this finally gives us the General Relativistic result

$$\Delta\phi = \frac{4M}{r_{\text{min}}} \equiv \frac{4GM}{c^2 r_{\text{min}}} = \frac{2R_{\text{S}}}{r_{\text{min}}} \tag{2.61}$$

This is exactly twice the deflection angle predicted by our earlier Newtonian treatment. If we take r_{min} to be the radius of the Sun (which would correspond to a light ray grazing the limb of the Sun from a background star observed during a total solar eclipse – see Figure 4) then we find that

$$\Delta\phi = \frac{4 \times 1.5 \times 10^3}{6.95 \times 10^8} = 8.62 \times 10^{-6} \text{ radians} \quad = \quad 1.77 \text{ arcsec} \tag{2.62}$$

The validity of the General Relativity result for the gravitational deflection of light was supported by the observations made by the Eddington expedition, and more recently has been repeatedly verified with much greater precision by radio observations of distant quasars when they are closely aligned with the solar limb.

The General Relativity light deflection formula also lies at the heart of the field of gravitational lensing – one of the most active areas of research in astronomy and cosmology today.

2.4.3 Gravitational lensing

Figure 5 shows the path of a light ray from a distant source deflected through an angle, α, by a close encounter with a point mass, M, which is exactly collinear with the source and observer. Since the deflection angle is small (in Figure 5 it is, of course, greatly exaggerated) we can represent the deflection as taking place at a single point, P, a perpendicular distance R_E from M, as shown. The photon path can then be approximated by the lines OP and PS.

By symmetry the point P must lie on circle of radius R_E in the plane perpendicular to OS, such that *any* point on the circle represents a possible point of deflection. Thus, the point mass, M, essentially acts as a **gravitational lens**, focussing light rays from the source at the observer, who will see an image of the source as a *ring*, or radius R_E.

This circle is shown schematically in the lower panel of Figure 5. It is known as an **Einstein Ring** and R_E is referred to as the **Einstein Radius** of the gravitational lens (although it was not Einstein but Chwolson, in 1924, who first predicted theoretically the existence of this ring).

We can determine the angular radius, θ_E of the Einstein Ring as follows. Let D_L and D_S

denote the distance of the gravitational lens and source respectively. From Figure 5 it is easy to see that

$$\theta_E + \beta = \alpha \tag{2.63}$$

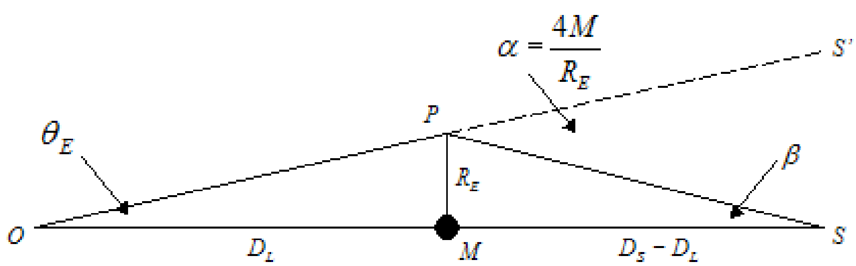
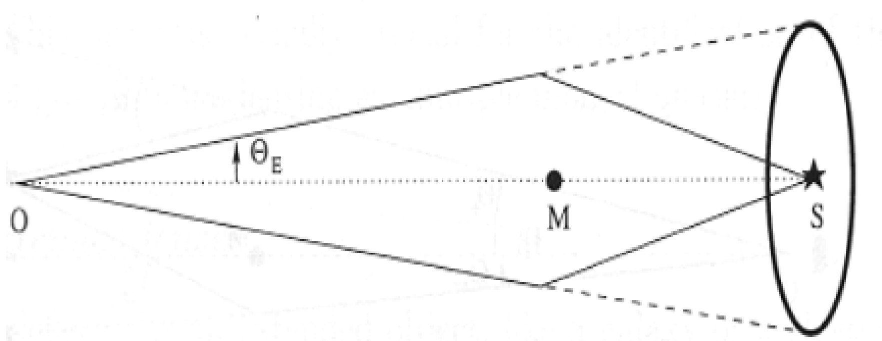


Figure 5

Gravitational deflection of light rays from a source at distance D_S, through an angle, α, by a point mass, *M*, at distance D_L from the observer, *O*. The source, deflecting mass and observer are assumed to be exactly collinear. For small deflection angle, α, we may regard the deflection as taking place at a single point, *P*, a distance R_E from *M*. The deflecting mass acts as a gravitational lens, producing an image of the source which is an **Einstein Ring** of angular radius θ_E, as shown below (from Mollerach & Roulet 2002)



Also,

$$\theta_E = \frac{R_E}{D_L} \tag{2.64}$$

and

$$\beta \;=\; \frac{R_E}{D_S - D_L} \;=\; \theta_E \frac{D_L}{D_S - D_L} \tag{2.65}$$

Substituting for β and α in equation (2.63) gives

$$\theta_E \,+\, \theta_E \frac{D_L}{D_S - D_L} \;=\; \theta_E \frac{D_S}{D_S - D_L} \;=\; \frac{4M}{R_E} \;=\; \frac{4M}{D_L \theta_E} \tag{2.66}$$

i.e.

$$\theta_E = \sqrt{\frac{4M(D_S - D_L)}{D_S D_L}} \tag{2.67}$$

Writing $x = D_L/D_S$ we can rewrite equation (2.67) as

$$\theta_E = \sqrt{\frac{4M(1-x)}{D_S\, x}} \tag{2.68}$$

Einstein Rings are also expected when the gravitational lens is an extended mass distribution which is spherically symmetric and is exactly aligned with the source and the observer; in this case M in equation (2.68) is the total mass inside the projected angular radius, θ_E. When the lensing mass is off-axis, and / or is not spherically symmetric then instead of an Einstein ring multiple images of the source are produced, the angular separation of which is of order θ_E.

Suppose, for example, a foreground galaxy lenses a background quasar or galaxy. It is straightforward to verify that equation (2.68) is conveniently expressed in this case as

$$\theta_E \simeq 3" \sqrt{\frac{M}{10^{12} M_\odot} \frac{10^9\,\mathrm{pc}}{D_S} \frac{(1-x)}{x}} \tag{2.69}$$

and consequently typical image separations are of order a few arcseconds, and should be resolvable separately. This regime is known as **strong gravitational lensing**.

In the past few decades hundreds of examples of strong lensing have been observed – mainly at radio and optical wavelengths – and the magnification and shape of the source images can be used to constrain the mass of the lens.

Consider now the case of lensing by stars within the Milky Way galaxy. In this case the angular Einstein Radius is conveniently expressed as

$$\theta_E \simeq 0.9\,\mathrm{mas}\sqrt{\frac{M}{M_\odot}\frac{10\,\mathrm{kpc}}{D_S}\frac{(1-x)}{x}} \tag{2.70}$$

(where 'mas' denotes milliarcseconds). Hence, the images in this case are (currently) too close together to be resolved. This regime is known as **gravitational microlensing**.

Although the microlensed images cannot be resolved, they do, however, change the apparent brightness of the source star. Moreover, if the lens is moving across the line of sight we can detect the change in magnification of the source as the projected lens-source angular separation changes.

Since the early 1990s a number of monitoring programs have been observing crowded stellar fields in the Galactic Bulge and the Magellanic Clouds, searching for the signatures of microlensing from intervening MACHOs (MAssive Compact Halo Objects). Several hundred microlensing events have been detected to date, and analysis of their **light curves** (i.e. how the apparent brightness of the sources changes with time) has allowed the mass of the lensing MACHOs to be constrained – results which have had important implications for cosmological searches for dark matter.

2.5 Gravitational redshift of light

In GR-I we used the Strong Principle of Equivalence to show that a photon 'climbing out' of a gravitational field is seen to be redshifted when it arrives at a distant observer. We now derive the same result within the framework of the Schwarzschild metric.

Suppose that light emitted at the event with coordinates (t_e, r_e) in the Schwarzschild metric travels along a radial null geodesic (i.e. with $d\theta = d\phi = 0$) to reach a distant observer at the event with coordinates (t_o, r_o). Then

$$ds^2 = 0 = -\left(1 - \frac{2M}{r}\right) dt^2 + \frac{dr^2}{1 - 2M/r} \tag{2.71}$$

i.e.

$$\int_{t_e}^{t_o} dt = t_o - t_e = \int_{r_e}^{r_o} \frac{dr}{1 - 2M/r} \tag{2.72}$$

Suppose we now think of the light as a wave, of frequency ν_e as measured in its rest frame. Suppose two adjacent wavecrests leave r_e at coordinate time t_e and $t_e + \Delta t_e$, and reach r_o at coordinate time t_o and $t_o + \Delta t_e$. Between the two emission events the elapsed proper time, $\Delta\tau_e$ is given by

$$\Delta\tau_e = \Delta t_e \sqrt{1 - \frac{2M}{r_e}} \equiv \frac{1}{\nu_e} \equiv \lambda_e \tag{2.73}$$

(recalling that $c = 1$).

For the observer at (t_o, r_o) the elapsed proper time between the arrival of the two wavecrests is

$$\Delta\tau_o = \Delta t_e \sqrt{1 - \frac{2M}{r_o}} \equiv \frac{1}{\nu_o} \equiv \lambda_o \tag{2.74}$$

Hence, the gravitational redshift of the light is given by

$$z \equiv \frac{\lambda_o - \lambda_e}{\lambda_e} = \sqrt{\frac{1 - 2M/r_o}{1 - 2M/r_e}} - 1 = \sqrt{\frac{r_e(r_o - R_S)}{r_o(r_e - R_S)}} - 1 \tag{2.75}$$

where R_S is the Schwarzschild radius of the central mass, M.

Consider light emitted from the Solar photosphere, at $r_e = R_\odot$, and observed at the distance of the Earth. Since $r_o >> r_e >> R_S$ (and ignoring the *blueshift* of the light as it falls into the gravity field of the Earth), equation (2.74) simplifies to

$$z \simeq \left(1 - \frac{M}{r_o}\right)\left(1 + \frac{M}{r_e}\right) - 1 \simeq \frac{M}{r_e} \tag{2.76}$$

Thus, the gravitational redshift of e.g. spectral line emitted at the Solar photosphere is

$$z = \frac{1.5 \times 10^3}{6.95 \times 10^8} \simeq 2 \times 10^{-6} \tag{2.77}$$

This effect is too small to be observed, since it is dwarfed in magnitude by the various astrophysical processes which contribute to the broadening of spectral lines (e.g. thermal, natural and collisional broadening). It has been successfully measured, however, in the spectra of **white dwarf stars**. These compact evolved stars have masses comparable to that of the Sun, but radii comparable to that of the Earth. Hence, one finds that

$$z = \frac{1.5 \times 10^3}{6.4 \times 10^6} \simeq 2.3 \times 10^{-4} \tag{2.78}$$

which is measurable (but only just!).

Perhaps the most convincing evidence for gravitational redshift comes from a terrestrial experiment, carried out by Pound, Rebka and Snider in 1960, using very high frequency gamma rays emitted at the foot and observed at the top of a 22m high tower at Harvard University. Although the predicted redshift was only about 2×10^{-15} in this case, it could be measured because the frequency of the gamma rays was known to extremely high precision, due to a resonance known as the **Mossbauer effect**. (See Green Schutz, page 120 for more details).

2.6 Gravitational time delay

The fourth classical test of General Relativity is closely related to the third test. In the previous section we saw that two events separated by coordinate time interval dt_e correspond to different intervals of proper time when observed at different radial coordinates, r_e and r_o respectively in the Schwarzschild metric.

In this section we show that, besides experiencing a deflection when passing close to a gravitating mass, light also experiences a time delay compared to the travel time in the absence of the mass.

The time delay has 2 contributions: the first is purely geometric, arising from the fact that the deflected trajectory is longer than the undeflected one. The second delay is gravitational, and is known as the **Shapiro Effect**. It comes about because clocks run more slowly in a gravitational field.

We can obtain a simple expression for the Shapiro Effect as follows. Consider first the invariant interval of the Schwarzschild metric expressed in its usual form

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dt^2 + \frac{dr^2}{\left(1 - \frac{2M}{r}\right)} + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2 \tag{2.79}$$

Suppose we now introduce a new radial coordinate, R, via the equation

$$r = R\left(1 + \frac{M}{2R}\right)^2 \tag{2.80}$$

(note that R still has dimensions of length). Since we are considering (for the moment at least) the GR effects of a weak gravitational field, such as that of the Sun, we may safely assume that $M << r$ and hence $M << R$. Thus, we may approximate equation (2.80) by

$$r \simeq R\left(1 + \frac{M}{R}\right) \tag{2.81}$$

Noting that

$$1 - \frac{2M}{r} = \frac{r - 2M}{r} \tag{2.82}$$

and substituting for r from equation (2.81), equation (2.79) becomes

$$ds^2 = -\left(\frac{1 - M/R}{1 + M/R}\right) dt^2 + \left(\frac{1 + M/R}{1 - M/R}\right) dr^2 + R^2 \left(1 + \frac{M}{R}\right)^2 \left[d\theta^2 + \sin^2\theta d\phi^2\right] \tag{2.83}$$

Using the binomial expansion for $x << 1$,

$$(1 + x)^n \simeq 1 + nx \tag{2.84}$$

and noting that to first order,

$$dr = dR \tag{2.85}$$

equation (2.83) simplifies further to

$$ds^2 = -\left(1 - \frac{2M}{R}\right) dt^2 + \left(1 + \frac{2M}{R}\right) \left[dR^2 + R^2 d\theta^2 + R^2 \sin^2\theta \, d\phi^2\right] \tag{2.86}$$

Defining Cartesian coordinates

$$X = R\sin\theta\cos\phi, \qquad Y = R\sin\theta\sin\phi, \qquad Z = R\cos\theta \tag{2.87}$$

and introducing the **weak field** (i.e. Newtonian) gravitational potential, ψ, given by

$$\psi = -\frac{M}{R} \equiv -\frac{GM}{R} \tag{2.88}$$

equation (2.86) reduces to

$$ds^2 = -(1 + 2\psi)\, dt^2 + (1 - 2\psi)\left[dX^2 + dY^2 + dZ^2\right] \tag{2.89}$$

Thus, we have reduced the spatial part of the metric to the Euclidean separation, re-scaled by the Newtonian potential.

Consider now the trajectory of a photon propagating between a source, at A, and an observer, at B, as shown in Figure 6. Since the deflection due to mass, M, is very small, we can approximate the trajectory as a straight line and introduce new coordinates, (x, y, z), such that the z-axis coincides with the photon trajectory. (Note that this means we neglect the geometric time delay). Consequently $dx = dy = 0$, and we may re-write equation (2.89) as

$$ds^2 = -(1 + 2\psi)\, dt^2 + (1 - 2\psi)\, dz^2 \tag{2.90}$$

Since $ds^2 = 0$, it follows that, to first order, the elapsed coordinate time between the emission and arrival of the photon is given by

$$dt^2 = \frac{1 - 2\psi}{1 + 2\psi} dz^2 \simeq (1 - 2\psi)^2\, dz^2 \tag{2.91}$$

i.e.

$$\int dt = \int (1 - 2\psi)\, dz \tag{2.92}$$

Hence,

$$t_{\mathrm{B}} - t_{\mathrm{A}} = (z_{\mathrm{B}} - z_{\mathrm{A}}) - 2 \int_{z_{\mathrm{A}}}^{z_{\mathrm{B}}} \psi(z) dz \tag{2.93}$$

The second term is the gravitational time delay:-

$$\delta t_{\mathrm{grav}} = -2 \int_{z_{\mathrm{A}}}^{z_{\mathrm{B}}} \psi(z) dz \tag{2.94}$$

or, in more conventional units

$$\delta t_{\mathrm{grav}} = -\frac{2}{c^3} \int_{z_{\mathrm{A}}}^{z_{\mathrm{B}}} \psi(z) dz \tag{2.95}$$

Thus, the gravitational time delay depends on the gravitational potential integrated along the photon's path.

The Shapiro delay has been measured using radar pulses bounced off Venus and Mercury, when those planets are closely aligned with the Sun. The results show impressive agreement with the predictions of General Relativity (see Figure 7).

The time delay is also very clearly seen in the arrival times of radio pulses from the binary pulsar system PSR 1913+16 (see Figure 8). Again the agreement with GR is excellent.

Finally, the Shapiro Effect is seen in the multiple images of gravitationally lensed quasars. It is possible to measure the time delay because quasars show intrinsic variations in their brightness and spectra. When such variations occur they are seen at different times in the different lensed images.

By carefully measuring the time delay between images and modelling the mass distribution of the lensing mass, it is possible to estimate the proper distance to the quasar. One may, then, combine this distance estimate with the observed redshift of the quasar to estimate the Hubble constant. This method has the advantage over more traditional 'Distance Ladder' estimates in that H_0 is measured on truly cosmological scales, where the observed redshift is unaffected by galaxy peculiar velocities.

2.6.1 Beyond the Shapiro effect?

The results derived in the previous section were for the case of a *static* metric. What if the gravitating mass is moving, or rotating, in the observer's frame as light propagates through the spacetime in its vicinity? In this case the Schwarzschild metric results are inapplicable because their assumption of a static central mass breaks down. In recent literature it has been suggested that observations of close binary pulsars may permit the measurement of gravitational time delay for the case where the gravitating mass *is* moving appreciably over the light crossing time of the system.

One exciting possibility is that such observations may permit measurement of the *speed of propagation* of gravitation. In General Relativity the gravitational field – which of course causes the deflection and time delay of a propagating light ray – is not generated instantaneously but propagates at the speed of light. Hence, crudely, one can reason that for a moving central mass, by the time the influence of the gravitational field has reached a material particle or photon in the surrounding spacetime, the mass will have moved to a new position – 'dragging' along with it the gravitational field which it generates.

Another (somewhat loose, but helpful) way of thinking about this is to say that the particle or photon experiences a 'retarded' gravitational field – arising from where the mass *was* a short time (in fact the light travel time, if GR is correct) before, rather than from where the mass has moved to by the time the effects of the gravitational field reach the particle. (A similar retarded phenomenon is seen with the electromagnetic fields generated by moving charges).

In recent literature there have been some claims that this gravitational retardation effect has been measured; these claims remain deeply controversial, but are likely to be a highly active area of theoretical and observational research in the future.

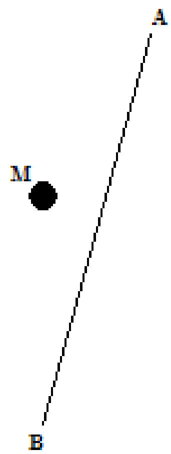
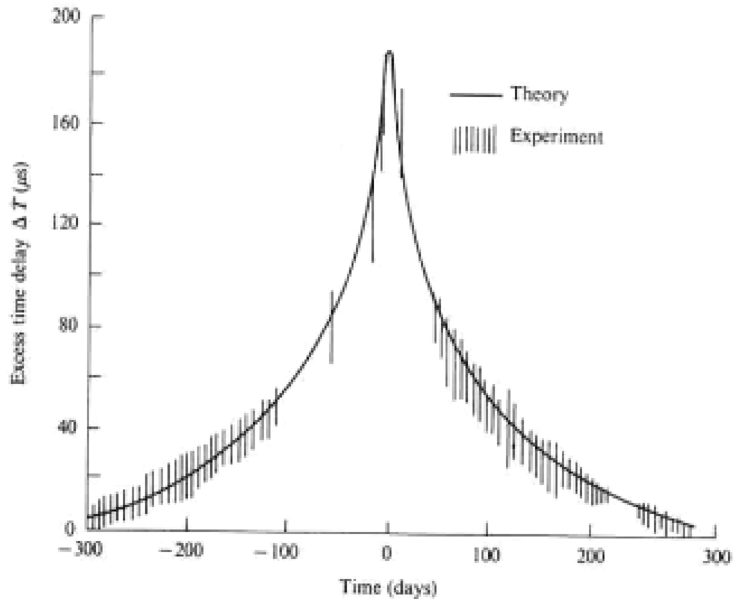


Figure 6

Trajectory of a light ray propagating from a source, A, to an observer, B, passing close to a mass, M

Figure 7

Radar time delay of signals reflected from Venus and passing close to the Sun; comparison of theory with observations. (From Shapiro,1970)



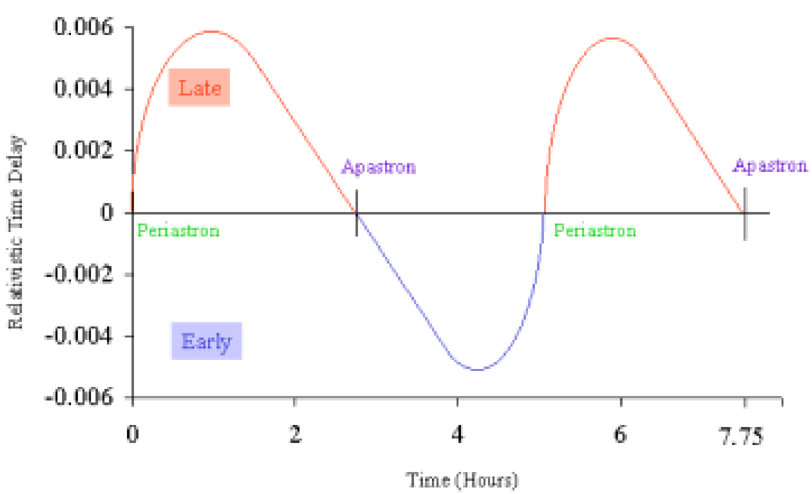


Figure 8

Time delay in the arrival of pulses from the binary pulsar PSR 1913+16. The observed data show excellent agreement with the predictions of GR

Chapter 3

Black Holes

3.1 Introduction

In Chapters 1 and 2 we first derived, and then investigated, the Schwarzschild solution for the static, spherically symmetric spacetime exterior to a star of (Newtonian) mass, M. We found that the line element for the Schwarzschild metric was given by

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dt^2 + \frac{dr^2}{\left(1 - \frac{2M}{r}\right)} + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2 \tag{3.1}$$

We noted in passing that this metric 'misbehaves' at $r = 2M$, since the g_{rr} component becomes infinite. However, since e.g. for a solar mass star, $M \simeq 1.5$km in geometrised units, it was clear that – for any 'normal' star – one would reach the physical surface of the star (where in any case one would have to replace the Schwarzschild solution for the exterior metric by the interior solution derived from solution of the Oppenheimer-Volkoff equation – which we do not consider in this course) before reaching the 'misbehaving' coordinate radius of $r = 2M$. It then follows from **Buchdahl's theorem** that, provided a star has finite central density and pressure, then the minimum allowed coordinate radius of its surface is still large enough to exceed the troublesome value of $r = 2M$ in the Schwarzschild metric.

3.2 Beyond white dwarfs and neutron stars?

Should we always assume a finite central density and pressure? In elementary astronomy classes we consider the internal structure of stellar remnants after they have left the Main Sequence. These stars are no longer supported by thermal pressure, since they are no longer producing heat from fusion reactions. Instead, they collapse and shrink under their own gravity until **quantum degeneracy pressure** becomes large enough to support them.

Simple arguments based on the **Heisenberg Uncertainty Principle** allowed us to estimate the maximum remnant mass which degeneracy pressure can support. For a **white dwarf** – where electron degeneracy pressure dominates – this maximum mass is $M \simeq 1.3 M_{\odot}$, and is known as the **Chandrasekhar Limit**. For a stellar remnant with a mass greater than the Chandrasekhar Limit, electron degeneracy pressure cannot support its weight and the collapse continues until **neutron degeneracy pressure** becomes large enough to halt it – we then have a **neutron star**. The astrophysical properties of white dwarfs and neutron stars will be left to other honours courses to discuss further. What interests us is simply the question: is there a maximum allowable neutron star mass?

For a stellar remnant of mass greater than about 2 solar masses, even neutron degeneracy pressure is insufficient; according to current physics, *nothing* can then halt the star's collapse under its own gravity, and the star shrinks to a singular point of infinite density – a **black hole**.

Does this singularity actually occur in nature? The current answer would have to be "nobody knows for sure". Many physicists believe that a (thus far incomplete) theory of quantum gravity would provide a new source of 'pressure' (from e.g. superstrings) to halt the star's collapse before reaching infinite density as $r \to 0$. Whether the central singularity *does* exist, at the very least it seems clear that the collapse of the stellar remnant can proceed beyond the point where the coordinate radius of the star satisfies $r = 2M$: the surface known as the **Schwarzschild radius**. We, therefore, need to consider carefully what exactly is going on in the Schwarzschild metric at $r = 2M$.

3.3 The nature of the Schwarzschild surface

Specifically, we want to know whether the misbehaviour of g_{rr} at $r = 2M$ is the result of a real, physical problem with the geometry of spacetime at the Schwarzschild radius, or is simply a consequence of our choice of coordinate system – what is referred to as a 'coordinate singularity'.

A more familiar example of a coordinate singularity is the North (or South) pole of the Earth. The poles are singular points in the latitude and longitude coordinate system because their longtiude is not uniquely defined. However, there is nothing physically different about these points (apart from being rather cold!) compared with any other on the Earth's surface.

We will determine the nature of the $r = 2M$ singularity by considering a material particle falling radially (i.e. with $d\theta = d\phi = 0$) towards the Schwarzschild radius from some finite coordinate radius, $R > 2M$. Suppose that the particle is released from rest at coordinate time $t = 0$ and proper time $\tau = 0$ in the particle's frame. Recall from equation (2.11) that

$$\left(\frac{dr}{d\tau}\right)^2 = k^2 - 1 - \frac{h^2}{r^2} + \frac{2M}{r}\left(1 + \frac{h^2}{r^2}\right) \tag{3.2}$$

Also, from equation (2.9)

$$\frac{d\phi}{d\tau} = \frac{h}{r^2} \tag{3.3}$$

Hence, for a radial trajectory, $h = 0$ and

$$\left(\frac{dr}{d\tau}\right)^2 = k^2 - 1 + \frac{2M}{r} \tag{3.4}$$

Since the particle is released from rest, it must follow that

$$k^2 - 1 = -\frac{2M}{R} \tag{3.5}$$

i.e.

$$\left(\frac{dr}{d\tau}\right)^2 = \frac{2M}{r} - \frac{2M}{R} \tag{3.6}$$

or, given that the particle is falling inwards

$$d\tau = -\frac{dr}{\sqrt{\frac{2M}{r} - \frac{2M}{R}}} \tag{3.7}$$

and the elapsed proper time experienced by the particle as it falls from $r = R$ to $r = 2M$ is given by

$$\Delta\tau = \int_{2M}^{R} \frac{dr}{\sqrt{\frac{2M}{r} - \frac{2M}{R}}} \tag{3.8}$$

We can see that this integral is clearly finite, since $2M/r$ tends to the (perfectly well behaved) value unity as $r \to 2M$. Hence, the particle reaches the Schwarzschild radius in a finite proper time.

What about the interval of coordinate time for the particle to reach the Schwarzschild radius? This follows from equation (2.8)

$$\frac{dt}{d\tau} = \frac{k}{1 - 2M/r} \quad = \quad \frac{\sqrt{1 - 2M/R}}{1 - 2M/r} \tag{3.9}$$

or, using, equation (3.7)

$$dt = -\frac{\left(\sqrt{1 - 2M/R}\right)\, dr}{\sqrt{\frac{2M}{r} - \frac{2M}{R}}\left(1 - \frac{2M}{r}\right)} \tag{3.10}$$

i.e.

$$\Delta t = \int_{2M}^{R} \frac{\left(\sqrt{1-2M/R}\right)\, dr}{\sqrt{\frac{2M}{r}-\frac{2M}{R}}\left(1-\frac{2M}{r}\right)} \tag{3.11}$$

This rather ugly-looking integral diverges as $r \to 2M$; i.e. the particle takes an infinite amount of coordinate time to reach the Schwarzschild radius. It is straightforward to show that the same result is true for a photon emitted radially at $r = R$; it appears, to a distant observer (whose clock essentially measures coordinate time) to take an infinite time to reach the Schwarzschild radius.

Thus, we have shown that the misbehaviour of the Schwarzschild metric at $r = 2M$ is only a coordinate singularity; the proper time experienced by a particle as it reaches the Schwarzschild radius is perfectly well-behaved, and the misbehaviour of g_{rr} in the Schwarzschild metric is purely a consequence of the coordinate system breaking down at $r = 2M$. Below we discuss briefly an alternative coordinate system in which the coordinate singularity at $r = 2M$ is removed.

What are the physical consequences for a material particle or photon crossing the Schwarzschild radius? Superficially, the answer is "nothing much". For example, astronauts free-falling in their spaceship towards the black hole would not be aware of any sudden, dramatic changes in the laws of physics as they reach $r = 2M$. In the astronaut's Local Inertial Frame, the results of any experiment would still agree with the predictions of Special Relativity – exactly as the strong equivalence principle requires[1]. Crossing the Schwarzschild radius has

[1]Depending on the *mass* of the black hole, however, the astronauts may have begun to experience the rather unpleasant phenomenon of 'spaghettification' by the time the spaceship crosses the Schwarzschild radius. This phenomenon arises because of the intense gradient in the gravity field of the black hole: if the astronauts are falling in feet first, then the gravitational pull is significantly different between their heads and feet. This

profound consequences for the ultimate fate of the astronaut, however, and is the reason why the $r = 2M$ surface is also known as the **Event Horizon** of the black hole.

Provided that the astronauts are at coordinate radius $r > 2M$, they can always fire their spaceship rockets and escape to infinity; as the spaceship approaches the Schwarzschild radius this would require ever-increasing amounts of fuel, but it remains possible in principle. As soon as the Schwarzschild radius is crossed, however, it is **impossible** for the spaceship to escape from the black hole's clutches. **All** trajectories must inevitably carry the spaceship to smaller coordinate radii – i.e. it cannot even remain at fixed r, no matter how much fuel it uses up trying to do so. The spaceship proceeds inexorably to $r = 0$ where (at least in classical General Relativity) there is a true, physical singularity of infinite density which will crush the astronauts and spaceship out of existence.

Once inside the Schwarzschild radius, the astronauts cannot even send a message – e.g. a single photon – to explain their fate to anxious friends waiting outside; even light emitted inside the Schwarzschild radius must follow a trajectory that inevitably leads inwards to the singularity[2].

tidal difference would stretch out the astronauts (and probably their spaceship too!) like a piece of spaghetti. However, this stretching is due to the tidal effect of the black hole's gravity field (which for a stellar mass black hole is very large for $r \sim 2M$) and does not directly have anything to do with crossing the Schwarzschild radius. Putting this another way, a LIF still exists at the Schwarzschild radius; it is simply that the LIF may be very small (and in particular much smaller than the astronauts!) due to the tidal gravitational field. For a supermassive black hole at the heart of a quasar, on the other hand, the Schwarzschild radius is much larger – e.g. if $M \simeq 10^6 M_\odot$, then $R_S = 3 \times 10^6$km. In this case the tidal stresses are small enough that the astronauts would cross the Schwarzschild radius without any noticeable discomfort.

[2]We will later see that, when quantum effects are included, it is not strictly true that light cannot escape from inside the Schwarzschild radius, but in purely classical terms this statement *is* true

The astronauts can, however, receive messages – and even food parcels! – from outside the Schwarzschild radius, although as seen by a distant observer they would appear to take an infinite time to reach the Schwarzschild radius. Thus, the astronauts are truly 'beyond the horizon' of any observer outside the Schwarzschild radius, which justifies the use of the terminology 'Event Horizon'.

3.4 Inside the event horizon

How can we prove these remarkable claims about crossing the Event Horizon? In fact they follow fairly straightforwardly from the Schwarzschild metric for $r < 2M$. (Although we have seen that a coordinate singularity exists at $r = 2M$, we can still use the Schwarzschild metric to describe the spacetime interior to the Event Horizon).

First, recall that the interval, ds^2, between any two neighbouring events – which we label (t, x, y, z) and $(t + dt, x + dx, y + dy, z + dz)$ in some coordinate system – in spacetime can be *null* (in which case $ds^2 = 0$), *spacelike*, ($ds^2 > 0$), or *timelike*, ($ds^2 < 0$). If the interval is spacelike, then one can find a Lorentz frame (S', say) in which the events occur at the same coordinate time – i.e. $dt = 0$ and $\sqrt{ds^2}$ is the proper distance between them. Thus, if $ds^2 > 0$, the two events cannot lie on the worldline of a material particle, since an observer in S' would then see the particle in two places at the same time – violating causality.

Now let us suppose that a particle is at rest inside the Event Horizon of a Schwarzschild black hole, and consider neighbouring events with (in spherical polar coordinates) $dr = d\theta = d\phi = 0$ and $dt \neq 0$. From equation (3.1),

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dt^2 \tag{3.12}$$

Since $r < 2M$, it follows that $ds^2 > 0$ – i.e. the interval between the events is **positive**, and so **cannot** lie on the worldline of the particle.

Thus, no particle can be stationary inside the Event Horizon. This is essentially because, for $r < 2M$ the roles of our coordinate labels r and t are in the opposite sense to their familiar interpretation far outside the Schwarzschild radius. Intervals with only $dt \neq 0$ behave as spacelike, instead of timelike, intervals, with the reverse being the case for intervals with only $dr \neq 0$. Consequently, just as in the exterior spacetime, far from the black hole, we can move freely through *space* but not through *time*, inside the Event Horizon we lose our ability to move freely in coordinate radius.

3.4.1 Removing the coordinate singularity

To proceed further we need to get round the problem of the misbehaving coordinate radius at $r = 2M$. We do this by introducing a new time coordinate, $\tilde{t}$, defined by[3]

$$\tilde{t} = t + 2M \ln \left| \frac{r}{2M} - 1 \right| \tag{3.13}$$

We can show that, in the coordinate system $(r, \tilde{t})$, and for simplicity taking $d\theta = d\phi = 0$,

$$ds^2 = -\left(1 - \frac{2M}{r}\right) d\tilde{t}^2 + \frac{4M}{r} dr d\tilde{t} + \left(1 + \frac{2M}{r}\right) dr^2 \tag{3.14}$$

which has **no** coordinate singularity at $r = 2M$. Note, however, that there is *still* a singularity at $r = 0$, as there should be as this is a true, physical singularity. A more general coordinate transformation which is similarly well-behaved at $r = 2M$ is discussed in some detail in Green Schutz, pg 292: it is known as **Kruskal – Szekeres** coordinates. We do not consider Kruskal – Szekeres coordinates further here, however.

[3]Remember that the coordinate time is only a convenient label and not of itself a meaningful, invariant physical quantity, so we are not altering any fundamental physical property of the spacetime geometry – only providing an easier way to visualise it

We obtain the equations of the **null cones** – i.e. the trajectories in spacetime which are the geodesics of photons – by setting $ds^2 = 0$ in equation (3.14), dividing through by dr^2 and solving for $d\tilde{t}/dr$. This gives us a quadratic equation, with roots

$$\frac{d\tilde{t}}{dr} = -1 \qquad \text{or} \qquad \frac{d\tilde{t}}{dr} = \frac{1 + 2M/r}{1 - 2M/r} \tag{3.15}$$

Figure 9 (following Berry – Principles of Cosmology and Gravitation) shows how these null cones change with coordinate radius. For $r >> R_S$, the null cones have an approximate slope of ± 1, i.e. as for flat Minkowski spacetime. As r approaches R_S, however, the null cones begin to 'tip over' and lean inwards. For $r = R_S$ one side of the null cone is vertical, while the other side (as at all other coordinate radii) has slope of exactly -1. Hence, at $r = R_S$ the timelike geodesics (which are the possible worldlines of material particles, and lie within the forward null cone) all point inwards towards $r = 0$. Moreover, for $r < R_S$ the null cones continue to tip over, so that even photons are inevitably carried to smaller coordinate radius. Thus, both material particles *and* photons, once inside the Event Horizon, **cannot** escape again, nor indeed can they avoid the fate of being pulled towards the singularity at $r = 0$. We can see from Figure 10, however, that our intrepid astronauts could indeed still receive messages and food parcels as they head towards their doom, inside the Event Horizon.

Figure 9

Null cones close to the Event Horizon of
a Schwarzschild black hole

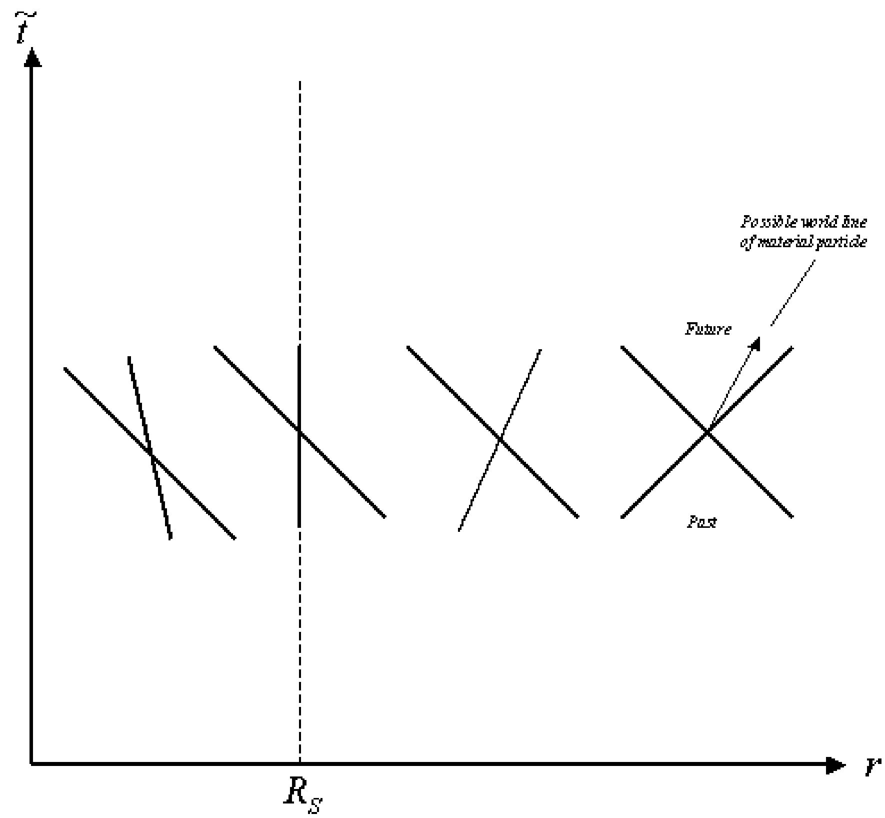
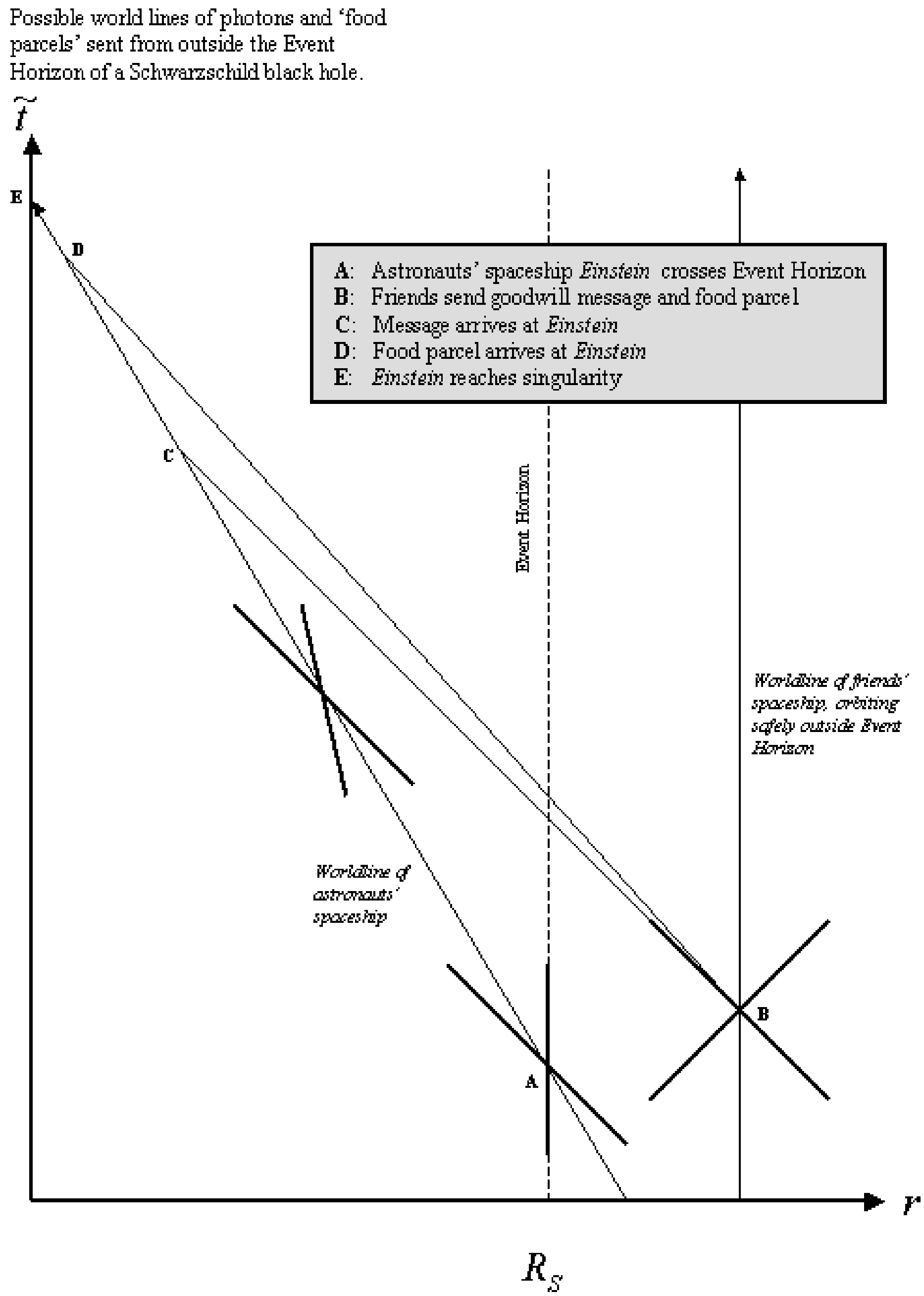


Figure 10



Possible world lines of photons and 'food parcels' sent from outside the Event Horizon of a Schwarzschild black hole.
$\tilde{t}$
E
D
C
A: Astronauts' spaceship *Einstein* crosses Event Horizon
B: Friends send goodwill message and food parcel
C: Message arrives at *Einstein*
D: Food parcel arrives at *Einstein*
E: *Einstein* reaches singularity
Event Horizon
Worldline of friends' spaceship, orbiting safely outside Event Horizon
Worldline of astronauts' spaceship
B
A
r
R_S

3.5 How black are black holes?

Since the region inside the Event Horizon of a Schwarzschild black hole is cut off from the outside Universe, we see the justification of the term 'hole': the interior of the Event Horizon really is like a 'hole' in spacetime. Is the name 'black hole' also justified, however? Does a collapsing simply star 'switch off' and become completely black?

At first we might think that the black hole would *not* switch off in this manner. We saw earlier that a material particle infalling radially towards the Schwarzschild radius takes an infinite amount of coordinate time to get there. Doesn't this mean, therefore, that as a massive star undergoes spherical collapse, a distant observer would see the light emitted from the star just before it collapses through the Schwarzschild radius, forever 'frozen' at the Event Horizon?

An answer to this question is provided when we realise that the light from the collapsing star is also redshifted, as it 'climbs out' of the star's gravity field. In Chapter 3 we derived equation (2.75) for the redshift, z, of light emitted at coordinate radius, r_e, and observed at coordinate radius, r_o.

$$z \equiv \frac{\lambda_o - \lambda_e}{\lambda_e} = \sqrt{\frac{1 - 2M/r_o}{1 - 2M/r_e}} - 1 = \sqrt{\frac{r_e(r_o - R_S)}{r_o(r_e - R_S)}} - 1 \qquad (3.16)$$

(Note that $z \to \infty$ as $r_e \to R_S$). Consider the bolometric luminosity of the star as it collapses, compared with the (constant) luminosity, L_C, which the star *would* have if we ignored General Relativistic effects. Suppose the light from the star is emitted radially at coordinates (t_e, r_e), and is then observed by a distant observer who is at coordinates (t_o, r_o). Then

$$L(t_o) = \frac{L_C}{(1+z)^2} \qquad (3.17)$$

We can understand equation (3.17) as follows. Firstly the energy of each photon received by the observer is redshifted by a factor of $(1+z)$; secondly the arrival times of the photons are also increased by the same factor, so that the luminosity, L, is reduced by a factor of $(1+z)^2$.

The light ray, being a null geodesic, satisfies

$$\int_{t_e}^{t_o} dt \quad = \quad \int_{r_e}^{r_o} \frac{dr}{1 - 2M/r} \quad \equiv \quad \int_{r_e}^{r_o} \frac{dr}{1 - R_S/r} \tag{3.18}$$

It is easy to verify that the integrand of the right hand side has integral $r + R_S \ln(r - R_S)$, so that

$$t_o - t_e \quad = \quad r_o - r_e + R_S \ln\left(\frac{r_o - R_S}{r_e - R_S}\right) \tag{3.19}$$

Taking $t_e = 0$ and re-arranging

$$\ln\left(\frac{r_e - R_S}{r_o - R_S}\right) \quad = \quad -\left[\frac{t_o - (r_o - r_e)}{R_S}\right] \tag{3.20}$$

i.e.

$$\frac{r_e - R_S}{r_o - R_S} \quad \propto \quad \exp\left[-t_o/R_S\right] \tag{3.21}$$

However, from equation (3.16)

$$\frac{r_e - R_S}{r_o - R_S} \quad = \quad \frac{1}{(1+z)^2} \tag{3.22}$$

so, from equation (3.17), it follows that

$$\frac{L(t_o)}{L_C} \quad \propto \quad \exp\left[-t_o/R_S\right] \tag{3.23}$$

or, re-introducing the speed of light, c

$$\frac{L(t_o)}{L_C} \quad \propto \quad \exp\left[-ct_o/R_S\right] \tag{3.24}$$

Thus, we see that the luminosity of the star falls off exponentially, and on a very short timescale – equal to the time for light to cross the Schwarzschild radius of the collapsing star[4].

We can express the dimming of the star in magnitudes, remembering that

$$M_1 - M_2 \quad = \quad -2.5 \log_{10} \frac{L_1}{L_2} \quad = \quad -\frac{2.5}{\ln 10} \ln \frac{L_1}{L_2} \quad \simeq \quad -1.09 \ln \frac{L_1}{L_2} \tag{3.25}$$

i.e.

$$M(t_o) - M_C \quad = \quad m(t_o) - m_C \quad \simeq \quad 1.09 \frac{ct_o}{R_S} \tag{3.26}$$

or

$$t_o \quad \simeq \quad \frac{R_S \left[m(t_o) - m_C \right]}{1.09c} \tag{3.27}$$

Suppose, for example, a 10 solar mass star, initially of apparent magnitude $m_C = 0$, undergoes core collapse. For this star, $R_S \simeq 30$km, so from equation (3.27) the star will have faded to an apparent magnitude of 28 (the limiting magnitude of the Hubble Space Telescope) in a time $t_o \simeq 2.6 \times 10^{-3}$ seconds.

Thus the redshifting the star's radiation means that effectively it *does* 'switch off' almost instantaneously as soon as it collapses down to its Schwarzschild radius, and the name 'black hole' is indeed an apt one.

3.6 Rotating black holes

So far in this chapter we have considered only static black holes. The treatment of more general situations is considerably more difficult, and in fact can usually only be handled numerically for specific cases. However, we will now briefly discuss one particular example of a more general black hole for which analytic results have been derived: a **Kerr black hole**.

[4] Our analysis here is a little simplistic, since the Schwarzschild metric is a *static* metric, and yet we are considering the behaviour of the metric as the star itself is collapsing. However, a more rigorous treatment would essentially give us the same result: that the luminosity of the collapsing star decreases exponentially

If the collapse of a star is nearly spherical then it can be shown that all of the non-spherical parts of the star's mass-energy distribution are radiated away as gravitational waves – except for some **angular momentum**. This residual angular momentum results in a **rotating black hole**, which can be described by the **Kerr metric**.

3.6.1 The Kerr metric

This metric is characterised by two constants, M and J, which we can identify by requiring that in the weak-field limit the behaviour of a test particle reduces to the predictions of Newtonian dynamics: M is the Newtonian mass of the star and J is the magnitude of its total angular momentum. Writing $a \equiv J/M$, the line element for the Kerr metric takes the (rather complicated) form

$$ds^2 = -\frac{\Delta - a^2 \sin^2\theta}{\rho^2}\,dt^2 - 4a\frac{Mr\sin^2\theta}{\rho^2}\,dt d\phi + \frac{\left(r^2+a^2\right)^2 - a^2\Delta\sin^2\theta}{\rho^2}\sin^2\theta\,d\phi^2 + \frac{\rho^2}{\Delta}\,dr^2 + \rho^2\,d\theta^2 \tag{3.28}$$

where

$$\Delta = r^2 - 2Mr + a^2 \tag{3.29}$$

and

$$\rho^2 = r^2 + a^2\cos^2\theta \tag{3.30}$$

A detailed discussion of the Kerr metric lies well beyond the scope of this course, but we note that the metric is **not** diagonal, because $g_{t\phi} \neq 0$. The presence of this term produces a quite remarkable effect known as 'frame dragging'. Before we can demonstrate this effect, however, we first consider an important result on conserved quantities along geodesics.

3.6.2 Conservation of four momentum along geodesics

Recall that the geodesic equation for a material particle

$$\frac{dv^\alpha}{d\tau} + \Gamma^\alpha_{\beta\delta} v^\beta v^\delta = 0 \tag{3.31}$$

We can rewrite this equation as

$$\left(\frac{\partial v^\alpha}{\partial x^\beta} v^\beta + \Gamma^\alpha_{\beta\delta} v^\beta v^\delta \right) = 0 \tag{3.32}$$

or

$$v^\beta \, v^\alpha{}_{;\beta} = 0 \tag{3.33}$$

(This equation basically *defines* a geodesic as a curve along which the tangent vector to the curve is parallel-transported.)

We can obtain another version of equation (3.33) in covariant form using

$$v^\alpha = g^{\alpha\sigma} v_\sigma \tag{3.34}$$

from which it is straightforward to show that

$$v^\alpha \, v_{\beta;\alpha} = 0 \tag{3.35}$$

If we introduce the contravariant and covariant components of the **four momentum**, defined by

$$p^\alpha = m v^\alpha \qquad \text{and} \qquad p_\alpha = m v_\alpha \tag{3.36}$$

where m is the rest mass of the particle, then it follows from equation (3.35) that

$$p^\alpha \, p_{\beta;\alpha} = 0 \tag{3.37}$$

or

$$p^\alpha \, p_{\beta,\alpha} \quad = \quad \Gamma^\gamma_{\beta\alpha} p^\alpha p_\gamma \quad = \quad \frac{1}{2} g^{\gamma\nu} \left(g_{\nu\beta,\alpha} + g_{\nu\alpha,\beta} - g_{\alpha\beta,\nu} \right) p^\alpha p_\gamma \tag{3.38}$$

After contraction and some index permutation this reduces to

$$p^{\alpha}\,p_{\beta,\alpha} \quad = \quad \frac{1}{2}g_{\nu\alpha,\beta}\,p^{\nu}p^{\alpha} \tag{3.39}$$

or

$$mv^{\alpha}\frac{\partial p_{\beta}}{\partial x^{\alpha}} \;=\; m\frac{dx^{\alpha}}{d\tau}\frac{\partial p_{\beta}}{\partial x^{\alpha}} \;=\; m\frac{dp_{\beta}}{d\tau} \;=\; \frac{1}{2}g_{\nu\alpha,\beta}\,p^{\nu}p^{\alpha} \tag{3.40}$$

We can see an important result from equation (3.40). If all the components of the metric are independent of coordinate x^{β}, then the right hand side of the equation is zero, which then implies that p_{β} is a constant along the geodesic.

3.6.3 Frame dragging in the Kerr Metric

If we look again at the Kerr metric interval, equation (3.28), we see that the components are independent of ϕ. The results of the previous section then imply that a material particle, moving along a geodesic in the Kerr metric, conserves the p_{ϕ} component of its four momentum.

The contravariant component, p^{ϕ}, is given by

$$p^{\phi} = g^{\phi\alpha}p_{\alpha} = g^{\phi\phi}p_{\phi} + g^{\phi t}p_{t} \tag{3.41}$$

The second term on the right hand side would be zero for an orthogonal metric, but the Kerr metric is *not* orthogonal. Similarly, for p^{t},

$$p^{t} = g^{t\alpha}p_{\alpha} = g^{tt}p_{t} + g^{t\phi}p_{\phi} \tag{3.42}$$

Consider now a material particle with zero angular momentum – i.e. with $p_{\phi} = 0$. Now

$$p^{t} = m\frac{dr}{d\tau} \qquad \text{and} \qquad p^{\phi} = m\frac{d\phi}{d\tau} \tag{3.43}$$

so that

$$\frac{d\phi}{dt} = \frac{p^\phi}{p^t} = \frac{g^{\phi t}}{g^{tt}} \neq 0 \tag{3.44}$$

However, $d\phi/dt$ defines the **angular velocity** of our material particle, as measured by a distant observer. Thus, we have the remarkable result that a zero angular momentum particle, dropped in free fall (i.e. in a locally inertial frame) radially inwards from infinity onto a rotating black hole, nevertheless acquires a non-zero angular velocity.

What is going on here? We can think of the intense gravity of the black hole 'dragging' the particle around in the same sense as its own rotation. This effect is known as the **dragging of inertial frames**. It will also occur for the metric of *any* rotating mass, although in most situations (e.g. the spacetime in the vicinity of the Earth) it is a tiny effect. Nevertheless, a satellite mission – **Gravity Probe B** – was recently launched by Stanford University to measure frame dragging from low Earth orbit. For more details see `http://einstein.stanford.edu`.

3.7 Hawking radiation

In 1974 the Cambridge physicist Stephen Hawking proved the startling result that black holes are not completely black! When one includes the effects of quantum mechanics, applied to electromagnetic fields near to a black hole, the classical General Relativistic result that *nothing* – not even light – can escape from inside the Event Horizon is no longer true: black holes leak photons continuously in a process that is now known as **Hawking radiation**.

To derive Hawking's result rigorously requires quantum field theory. However, we can get a feel for the main features of Hawking radiation by applying the Heisenberg Uncertainty Principle: whereas up until now we have treated photons (and indeed material particles) as precise points along a trajectory in spacetime, the Uncertainty Principle implies that, in fact, photons and material particles are not 'localisable' – their position and velocity cannot be determined to arbitrary precision.

One form of the Uncertainty Principle states that

$$\Delta E \Delta t \sim \hbar = \frac{h}{2\pi} \tag{3.45}$$

where ΔE is the minimum uncertainty in the energy of a particle which exists in a particular quantum state for time Δt. According to quantum field theory, the vacuum of empty space is not empty at all, but is filled with quantum fluctuations in the fields which it contains. For electromagnetic fields, these fluctuations consist of pairs of 'virtual photons', the existence of which violates energy conservation, but provided the photons exist for less than $\Delta t \simeq \Delta E/\hbar$ this is permissible; the energy which they borrow from the vacuum is effectively 'repaid' before the vacuum has noticed.

Even close to the Event Horizon of a black hole, spacetime is locally flat and will therefore be filled with these quantum fluctuations. Consider a fluctuation which produces two photons: one of energy, E, and the other of energy, $-E$. In the flat spacetime of Special Relativity, the negative energy photon cannot propagate freely, since it would be moving backwards in time (see Green Schutz, pgs 303-304), and so it must recombine with the positive energy photon within a time $\Delta t \leq \hbar/E$.

Suppose, however, the pair of virtual photons is produced in the locally flat spacetime just outside the Event Horizon of a black hole. The negative energy member of the pair therefore has a chance of crossing the horizon before a time $\hbar/E$ has elapsed. Once inside the horizon, the negative energy photon *can* propagate freely because the timelike and spacelike roles of the t and r coordinates are interchanged. The negative energy photon **must** move radially inwards, and therefore cannot escape again from the Event Horizon, but that still leaves the positive energy photon outside the horizon, free to propagate and to escape to infinity.

Suppose we measure the energy, $\mathcal{E}$, of the photon in a locally inertial frame just outside the Event Horizon, at coordinate radius $r = 2M + \epsilon$. Any particle in this locally inertial frame will begin free-falling towards the Event Horizon, reaching it after a proper time, $\Delta\tau$, given by, from equation (3.8)

$$\Delta\tau = \int_{2M+\epsilon}^{2M} \frac{dr}{\sqrt{\frac{2M}{r} - \frac{2M}{2M+\epsilon}}} \tag{3.46}$$

For $\epsilon << 2M$ this reduces to

$$\Delta\tau = 2\,(2M\epsilon)^{1/2} \tag{3.47}$$

If we set $\mathcal{E} = \Delta E$ and $\Delta\tau = \Delta t$ in the Uncertainty Principle, we find that

$$\mathcal{E} = \frac{1}{2}\hbar\,(2M\epsilon)^{-1/2} \tag{3.48}$$

The energy of the outgoing photon when it reaches infinity can be shown to be

$$E_\infty = \mathcal{E}\left(\frac{\epsilon}{2M}\right)^{1/2} = \frac{\hbar}{4M} = \frac{h}{8\pi M} \tag{3.49}$$

Hawking derived the more rigorous result that the outgoing photons have a **black-body** spectrum with temperature $T = \hbar/8\pi kM$ and typical energy $E = kT = \hbar/8\pi M$, which differs from our result only by a factor of 2π. (The fact that Hawking radiation is black-body makes sense, even without formal proof, since classically a black hole is a perfect absorber of radiation).

3.7.1 Lifetime of a black hole

We see that the temperature of a black hole is inversely proportional to its mass. For, e.g., a black hole of 10 solar masses, we find that

$$T = \frac{\hbar}{8\pi k M} = 1.5 \times 10^{-44}\,\mathrm{K} \tag{3.50}$$

which is an incredibly small, but still formally non-zero, temperature.

The luminosity of the Hawking radiation from the black hole is, according to the Stefan-Boltzmann law, proportional to AT^4, where A is the area of the Horizon. Since

$$A \quad = \quad 4\pi R^2 \quad = \quad 4\pi(2M)^2 \quad = \quad 16\pi M^2 \tag{3.51}$$

for a Schwarzschild black hole, this means that

$$L \quad \propto \quad M^{-2} \tag{3.52}$$

This luminosity must come at the expense of a decrease in the mass of the black hole (remember, the escape to infinity of the positive energy photon is accompanied by the capture of a *negative* energy photon, which reduces the mass of the black hole). Hence

$$\frac{dM}{dt} \quad \propto \quad M^{-2} \tag{3.53}$$

from which it follows that the lifetime, τ, of a black hole satisfies

$$\tau \quad \propto \quad M^3 \tag{3.54}$$

i.e. the more massive the black hole, the longer it lives, and the cooler its temperature. Determining the constant of proportionality, it is found that

$$\left(\frac{\tau}{10^{10}\,\mathrm{yr}}\right) \quad = \quad \left(\frac{M}{10^{12}\,\mathrm{kg}}\right)^3 \tag{3.55}$$

So for a stellar mass black hole, Hawking radiation is completely negligible. In some cosmological theories, however, **primordial black holes** can form in the very early Universe; equation (3.55) says that a primordial black hole of mass around 10^{12} kg would have 'evaporated' on a timescale comparable to the age of the Universe. Since the black-body temperature of the black hole increases as its mass decreases, the evaporation would end with a very short and violent burst of gamma rays. Detection of such a gamma ray background would provide a useful observational test for the existence of primordial black holes.

3.7.2 Unifying gravity and thermodynamics

Some years before his groundbreaking work on Hawking radiation, Stephen Hawking proved what is now known as **Hawking's area theorem**. This stated that, if treated purely classically, the area of the Event Horizon of a black hole can never decrease, i.e.

$$\frac{dA}{dt} \geq 0 \tag{3.56}$$

In 1973, shortly after the Area theorem was discovered, Jakob Bekenstein drew an analogy with thermodynamics: if one related the area of the Event Horizon to the entropy of the black hole, the Area theorem was equivalent to the **Second Law of Thermodynamics** – that entropy never decreases with time. The analogy seemed incomplete, however, since it was thought that black holes did not have a thermodynamic temperature. Remarkably, however, Hawking radiation provided a rationale for assigning just such a temperature.

For a Schwarzschild black hole, $A = 16\pi M^2$, so that

$$dA = 32\pi M dM \tag{3.57}$$

or

$$dM = \frac{1}{32\pi M} dA = \frac{\hbar}{8\pi k M} d\left(\frac{Ak}{4\hbar}\right) \tag{3.58}$$

where dM is the change in the total energy of the black hole. From the previous section, we can re-write this as

$$dE = T\,dS \tag{3.59}$$

In other words, a black hole behaves like a thermodynamic black-body with temperature, $T = \hbar/8\pi kM$ and entropy $S = Ak/4\hbar$. Although, as the black hole radiates, it loses mass and thus the area of its Event Horizon *does* decrease (Hawking's Area theorem only applies to a classical black hole) the entropy of the Hawking radiation which it produces more than compensates for this and the *net* entropy of the black hole and the exterior Universe does *not* reduce – in accordance with the Second Law of Thermodynamics.

Thus, the discovery of Hawking radiation led to a remarkable unification of gravitation and thermodynamics.

Chapter 4

Gravitational Radiation

4.1 Non-stationarity

In Chapters 1 and 2 we have considered a static metric, which we defined as a metric for which we can find a time coordinate, t, satisfying

1. all metric components are independent of t
2. the metric is unchanged if we apply the transformation $t \to -t$

A metric which satisfies property (1) but not property (2) is known as **stationary**. An example is the metric of a spherically symmetric star which is *rotating*: reversing the time coordinate changes the sense of the rotation, even though one can find a coordinate system in which the metric components are all independent of time. In the next chapter (time permitting) we will consider the stationary metric of a rotating black hole.

In this chapter we explore some consequences of also relaxing the assumption of property (1), by considering spacetimes in which the metric components are time dependent. This can happen when the source of the gravitational field is varying. One of the most important predictions of General Relativity was that such a time-varying gravitational field generates **gravitational radiation** – 'ripples' in spacetime, which propagate at the speed of light. These gravitational waves are vanishingly weak for all but the most extreme astrophysical situations, and their detection and measurement presents enormous technological challenges compared with electromagnetic radiation.

For several decades physicists at Glasgow University have been at the forefront of the global effort to build gravitational wave detectors. Several such detectors have recently gone 'online' and are now searching for gravitational wave signatures. Although no signals have been detected directly to date, the first detections are confidently expected within the next few years. Moreover, for several decades there has been strong *indirect* evidence for the existence of gravitational radiation because the waves carry away energy from the source of the gravitational field. Observations of the binary pulsar system PSR 1913+16 show that its orbital semi-major axis is 'shrinking' (see Figure 11) and the rate of decrease in the semi-major axis is in excellent agreement with the calculated prediction based on the energy loss from gravitational radiation.

Although the study of gravitational radiation is one of the most exciting and active fields of research in astrophysics and cosmology today, unfortunately it is also an extremely technically difficult subject and mainly lies well beyond the scope of this course. Nevertheless, in this chapter we will try to highlight some of the important characteristics of gravitational radiation, beginning with a discussion of how plane gravitational waves arise as the free-space solutions to Einstein's equations in the 'linearised theory' of a weak gravitational field.

4.2 Weak gravitational fields

4.2.1 'Nearly' flat spacetimes

Since spacetime is flat in the absence of a gravitational field, a weak gravitational field is one in which spacetime is 'nearly' flat. What we mean by 'nearly' here is that we can find a coordinate system in which the metric has components

$$g_{\alpha\beta} = \eta_{\alpha\beta} + h_{\alpha\beta} \tag{4.1}$$

where

$$\eta_{\alpha\beta} = \text{diag}\ (-1, 1, 1, 1) \tag{4.2}$$

is the Minkowski metric of Special Relativity, and $|h_{\alpha\beta}| << 1$ for all α and β.

A coordinate system which satisfies equations (4.1) and (4.2) is referred to as a 'Nearly Lorentz' coordinate system. Notice that we say that we can find *a* coordinate system satisfying these equations. It certainly does *not* follow that for *any* choice of coordinate system we can write the metric components of the nearly flat spacetime in the form of equations (4.1) and (4.2). Indeed, even if the spacetime is precisely Minkowskian, we could adopt (somewhat perversely perhaps) a coordinate system in which the metric components were very far from the simple form of equation (4.2). This does not affect the validity of the principle of general covariance in General Relativity: remember that if we write down a tensor equation in one coordinate system, the equation remains valid in any other coordinate system, but the tensor *components* change from one coordinate system to another.

In some coordinate systems, therefore, the components may be enormously more complicated than in others. The secret to *solving* tensor equations in General Relativity is, often, to first choose a coordinate system in which the components are as simple as possible. In that sense, equations (4.1) and (4.2) represent a 'good' choice of coordinate system; just as equation (4.2) represents the simplest form we can find for the metric components in flat spacetime, so equation (4.1) represents the metric components of a nearly flat spacetime in their simplest possible form.

The coordinate system in which one may express the metric components of a nearly flat spacetime in the form of equations (4.1) and (4.2) is certainly not unique. If we have identified such a coordinate system then we can find (an infinite family of) others by carrying out particular coordinate transformations. There are two types of coordinate transformations which preserve the properties of equations (4.1) and (4.2). These are known as **Background Lorentz transformations** and **Gauge transformations**.

4.2.2 Background Lorentz transformations

Suppose we are in the Minkowski spacetime of Special Relativity, and we define the inertial frame, S, with coordinates (t, x, y, z). Suppose we then transform to another inertial frame, S', corresponding to a 'Lorentz boost' of velocity v in the direction of the positive x-axis. Under the Lorentz transformation, S' has coordinates given by, in matrix form

$$(t', x', y', z')^T = \begin{pmatrix} \gamma & -v\gamma & 0 & 0 \\ -v\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} (t, x, y, z)^T \tag{4.3}$$

where $\gamma = \left(1 - v^2\right)^{-1/2}$. (Remember that we are taking $c = 1$). We can write this in more compact notation as

$$x'^{\alpha} = \Lambda^{\alpha'}_{\beta} x^{\beta} \quad \equiv \quad \frac{\partial x'^{\alpha}}{\partial x^{\beta}} x^{\beta} \tag{4.4}$$

The Lorentz matrix has inverse, corresponding to a boost of velocity v along the *negative* x-axis, given by

$$(t, x, y, z)^T = \begin{pmatrix} \gamma & v\gamma & 0 & 0 \\ v\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} (t', x', y', z') \tag{4.5}$$

or

$$x^{\alpha} = \Lambda^{\alpha}_{\beta'} x'^{\beta} \quad \equiv \quad \frac{\partial x^{\alpha}}{\partial x'^{\beta}} x'^{\beta} \tag{4.6}$$

Now suppose we are in a nearly flat spacetime in which we have identified nearly Lorentz coordinates (t, x, y, z) satisfying equations (4.1) and (4.2). Suppose we now transform to a new coordinate system (t', x', y', z') defined such that

$$x'^{\alpha} = \Lambda^{\alpha'}_{\beta} x^{\beta} \tag{4.7}$$

i.e. where the transformation matrix is identical in form to equation (4.3) for some constant v. In this new coordinate system the metric components take the form

$$g'_{\alpha\beta} \;=\; \Lambda^{\mu}_{\alpha'}\Lambda^{\nu}_{\beta'}g_{\mu\nu} \;=\; \frac{\partial x^{\mu}}{\partial x'^{\alpha}}\frac{\partial x^{\nu}}{\partial x'^{\beta}}g_{\mu\nu} \tag{4.8}$$

Substituting from equation (4.1) this becomes

$$g'_{\alpha\beta} \;=\; \frac{\partial x^{\mu}}{\partial x'^{\alpha}}\frac{\partial x^{\nu}}{\partial x'^{\beta}}\eta_{\mu\nu} + \frac{\partial x^{\mu}}{\partial x'^{\alpha}}\frac{\partial x^{\nu}}{\partial x'^{\beta}}h_{\mu\nu} \tag{4.9}$$

Because of the particular form of the coordinate transformation in this case, it follows that

$$g'_{\alpha\beta} \;=\; \eta'_{\alpha\beta} + \frac{\partial x^{\mu}}{\partial x'^{\alpha}}\frac{\partial x^{\nu}}{\partial x'^{\beta}}h_{\mu\nu} \;=\; \eta'_{\alpha\beta} + h'_{\alpha\beta} \tag{4.10}$$

Thus, provided we consider **only** transformations of the form of equation (4.3), the components of $h_{\mu\nu}$ transform **as if** they are the components of a $(0,2)$ tensor defined on a Background flat spacetime. Our original 'Nearly Lorentz' coordinate system remains 'Nearly Lorentz' in the new coordinate system. In other words, our 'nearly flat' spacetime still looks 'nearly flat' under the Background Lorentz transformation.

4.2.3 Gauge transformations

Suppose now we make a very small change in our coordinate system by applying a coordinate transformation of the form

$$x'^{\alpha} = x^{\alpha} + \xi^{\alpha}(x^{\beta}) \tag{4.11}$$

i.e. where the components ξ^{α} are functions of the coordinates, $\{x^{\alpha}\}$. It then follows that

$$\frac{\partial x'^{\alpha}}{\partial x^{\beta}} = \delta^{\alpha}_{\beta} + \xi^{\alpha}_{,\beta} \tag{4.12}$$

From equation (4.11) we can also write

$$x^{\alpha} = x'^{\alpha} - \xi^{\alpha}(x^{\beta}) \tag{4.13}$$

If we now demand that the ξ^{α} are small, in the sense that

$$|\xi^{\alpha}{}_{,\beta}| << 1 \quad \text{for all} \quad \alpha, \beta \tag{4.14}$$

then it follows by the chain rule that

$$\frac{\partial x^{\alpha}}{\partial x'^{\gamma}} = \delta^{\alpha}_{\gamma} - \frac{\partial x^{\beta}}{\partial x'^{\gamma}}\frac{\partial \xi^{\alpha}}{\partial x^{\beta}} \simeq \delta^{\alpha}_{\gamma} - \xi^{\alpha}_{,\gamma} \tag{4.15}$$

where we have neglected terms higher than first order in small quantities. We have also used the fact that the components of the Kronecker delta are the same in any coordinate system.

Suppose now that the unprimed coordinate system is nearly Lorentz – i.e. the metric components satisfy equations (4.1) and (4.2). What about the metric components in the primed coordinate system?

Since the metric is a tensor, we know that

$$g'_{\alpha\beta} = \frac{\partial x^\mu}{\partial x'^\alpha} \frac{\partial x^\nu}{\partial x'^\beta} g_{\mu\nu} \tag{4.16}$$

Substituting from equations (4.1) and (4.15) this becomes, to first order

$$g'_{\alpha\beta} = \left(\delta^\mu_\alpha \delta^\nu_\beta - \xi^\mu_{,\alpha} \delta^\nu_\beta - \xi^\nu_{,\beta} \delta^\mu_\alpha \right) \eta_{\mu\nu} + \delta^\mu_\alpha \delta^\nu_\beta h_{\mu\nu} \tag{4.17}$$

This simplifies to

$$g'_{\alpha\beta} = \eta_{\alpha\beta} + h_{\alpha\beta} - \xi_{\alpha,\beta} - \xi_{\beta,\alpha} \tag{4.18}$$

Note that in equation (4.18) we have defined

$$\xi_\alpha = \eta_{\alpha\nu} \xi^\nu \tag{4.19}$$

and we have also used the fact that all the partial derivatives of $\eta_{\alpha\nu}$ are zero.

Thus, equation (4.19) has the same form as equation (4.1) provided

$$h'_{\alpha\beta} = h_{\alpha\beta} - \xi_{\alpha,\beta} - \xi_{\beta,\alpha} \tag{4.20}$$

Note that if $|\xi^{\alpha}{}_{,\beta}|$ are small, then so too are $|\xi_{\alpha,\beta}|$, and hence $h'_{\alpha\beta}$. Thus, our new primed coordinate system is *still* nearly Lorentz.

The above results tell us that – given that we have identified a coordinate system which is nearly Lorentz – we can add an arbitrary small 'vector' ξ^{α} to the coordinates, x^{α}, without altering the validity of our assumption that spacetime is nearly flat. We can, therefore, choose the components ξ^{α} to make Einstein's equations as simple as possible. We call this step choosing a **gauge** for the problem – a name which has resonance with a similar procedure in electromagnetism – and the coordinate transformation given by equation (4.20) is known as a **gauge transformation**.

4.3 Einstein's equations for a weak gravitational field

If we can work in a nearly Lorentz coordinate system for a nearly flat spacetime this simplifies Einstein's equations considerably, and leads us to spot that the deviations from the metric of Minkowski spacetime – the components $h_{\alpha\beta}$ in equation (4.1) – obey a **wave equation**.

Before we arrive at this key result, however, we have some algebraic work to do first. We begin by deriving an expression for the Riemann-Christoffel tensor in a weak gravitational field.

4.3.1 Riemann-Christoffel tensor for a weak gravitational field

In its fully covariant form the Riemann-Christoffel tensor is given by

$$R_{\alpha\beta\gamma\delta} = g_{\alpha\mu} R^{\mu}_{\beta\gamma\delta} = g_{\alpha\mu} \left[\Gamma^{\sigma}_{\beta\delta} \Gamma^{\mu}_{\sigma\gamma} - \Gamma^{\sigma}_{\beta\gamma} \Gamma^{\mu}_{\sigma\delta} + \Gamma^{\mu}_{\beta\delta,\gamma} - \Gamma^{\mu}_{\beta\gamma,\delta} \right] \tag{4.21}$$

Recall from the previous section that, if we are considering Background Lorentz transformations – i.e. if we restrict our attention only to the class of coordinate transformations which obey equation (4.7) – then the metric perturbations, $h_{\alpha\beta}$, transform as if they are the components of a $(0, 2)$ tensor defined on flat, Minkowski spacetime. In this case the Christoffel symbols of the first two bracketed terms on the right hand side of equation (4.21) are equal to zero. It is then easy to show that, to first order in small quantities, the Riemann-Christoffel tensor reduces to

$$R_{\alpha\beta\gamma\delta} = \frac{1}{2} \left(h_{\alpha\delta,\beta\gamma} + h_{\beta\gamma,\alpha\delta} - h_{\alpha\gamma,\beta\delta} - h_{\beta\delta,\alpha\gamma} \right) \tag{4.22}$$

Moreover, it can also be shown that, to first order, equation (4.22) is invariant under gauge transformations – i.e. the components of the Riemann-Christoffel tensor are independent of the choice of gauge.

4.3.2 Einstein's tensor for a weak gravitational field

From equations (4.21) and (4.22) we can contract the Riemann-Christoffel tensor and thus obtain an expression for the Ricci tensor in linearised form. This can be shown to take the form

$$R_{\mu\nu} = \frac{1}{2}\left(h^{\alpha}_{\mu,\nu\alpha} + h^{\alpha}_{\nu,\mu\alpha} - h_{\mu\nu,\alpha}{}^{,\alpha} - h_{,\mu\nu}\right) \tag{4.23}$$

where we have written

$$h \equiv h^{\alpha}_{\alpha} = \eta^{\alpha\beta} h_{\alpha\beta} \tag{4.24}$$

Note that we have raised the indices of the components $h_{\alpha\beta}$ using $\eta^{\alpha\beta}$ since $h_{\alpha\beta}$ behaves like a $(0,2)$ tensor defined on a flat spacetime, for which the metric is $\eta^{\alpha\beta}$. The derivation of equation (4.23) also uses the fact that all partial derivatives of $\eta^{\alpha\nu}$ are zero.

Note also, that we have introduced the notation, generalising the definition of equation (4.19)

$$f^{\alpha} = \eta^{\alpha\nu} f_{\nu} \tag{4.25}$$

where f^{α} are the components of a 'vector'. We can also extend this notation for raising and lowering indices to the components of more general geometrical objects, and to their partial derivatives. For example, in equation (4.23)

$$h_{\mu\nu,\alpha}{}^{,\alpha} \;=\; \eta^{\alpha\sigma}\left(h_{\mu\nu,\alpha}\right)_{\sigma} \;=\; \eta^{\alpha\sigma}\, h_{\mu\nu,\alpha\sigma} \tag{4.26}$$

After a further contraction of the Ricci tensor, to obtain the curvature scalar, R, where

$$R = \eta^{\alpha\beta} R_{\alpha\beta} \tag{4.27}$$

and substitution into the equation

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu} R \tag{4.28}$$

we obtain, after considerable further algebraic manipulation, an expression for the Einstein tensor, $G_{\mu\nu}$, in linearised, fully covariant form

$$G_{\mu\nu} = \frac{1}{2}\left[h_{\mu\alpha,\nu}{}^{,\alpha} + h_{\nu\alpha,\mu}{}^{,\alpha} - h_{\mu\nu,\alpha}{}^{,\alpha} - h_{,\mu\nu} - \eta_{\mu\nu}\left(h_{\alpha\beta}{}^{,\alpha\beta} - h_{,\beta}{}^{,\beta}\right)\right] \tag{4.29}$$

This rather messy expression can be simplified a little by introducing a new form (rather like a 're-scaling') for the metric perturbations

$$\overline{h}_{\mu\nu} \equiv h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu} h \tag{4.30}$$

after which equation (4.29) becomes

$$G_{\mu\nu} = -\frac{1}{2}\left[\overline{h}_{\mu\nu,\alpha}{}^{,\alpha} + \eta_{\mu\nu}\overline{h}_{\alpha\beta}{}^{,\alpha\beta} - \overline{h}_{\mu\alpha,\nu}{}^{,\alpha} - \overline{h}_{\nu\alpha,\mu}{}^{,\alpha}\right] \tag{4.31}$$

4.3.3 Einstein's equations for a weak gravitational field

Having ploughed our way through all of the above algebra, we can now write down Einstein's equations in their linearised, fully covariant form for a weak gravitational field, in terms of the (re-scaled) metric perturbations, $\overline{h}_{\mu\nu}$. Since

$$G_{\mu\nu} = 8\pi T_{\mu\nu} \tag{4.32}$$

it follows that

$$-\overline{h}_{\mu\nu,\alpha}{}^{,\alpha} - \eta_{\mu\nu}\overline{h}_{\alpha\beta}{}^{,\alpha\beta} + \overline{h}_{\mu\alpha,\nu}{}^{,\alpha} + \overline{h}_{\nu\alpha,\mu}{}^{,\alpha} = 16\pi T_{\mu\nu} \tag{4.33}$$

It can be shown that we can *always* find a gauge transformation which sets the last three terms on the left hand side of equation (4.33) equal to zero. This transformation is the **Lorentz gauge** referred to earlier, and is equivalent to adopting a coordinate system in which

$$\overline{h}^{\mu\alpha}{}_{,\alpha} = 0 \tag{4.34}$$

i.e. the **divergence** of the metric perturbations is equal to zero. Thus, in the Lorentz gauge, the linearised Einstein field equations reduce to the somewhat simpler form

$$-\overline{h}_{\mu\nu,\alpha}{}^{,\alpha} = 16\pi T_{\mu\nu} \tag{4.35}$$

4.3.4 Solution to Einstein's equations in free space

The free space solutions of equation (4.35) are solutions of the equation

$$\overline{h}_{\mu\nu,\alpha}{}^{,\alpha} = 0 \tag{4.36}$$

or, using equation (4.26)

$$\overline{h}_{\mu\nu,\alpha}{}^{,\alpha} \quad \equiv \quad \eta^{\alpha\alpha}\overline{h}_{\mu\nu,\alpha\alpha} \tag{4.37}$$

In fact, when we write out equation (4.37) explicitly, it takes the form

$$\left(-\frac{\partial^2}{\partial t^2} + \nabla^2 \right) \overline{h}_{\mu\nu} \quad = \quad 0 \tag{4.38}$$

Remembering that we are taking $c = 1$, if instead we write

$$\eta^{00} = -\frac{1}{c^2} \tag{4.39}$$

then equation (4.38) can be re-written as

$$\left(-\frac{\partial^2}{\partial t^2} + c^2 \nabla^2 \right) \overline{h}_{\mu\nu} \quad = \quad 0 \tag{4.40}$$

This is a key result. Equation (4.40) has the mathematical form of a wave equation, propagating with speed c. Thus, we have shown that the metric perturbations – the 'ripples' in spacetime produced by disturbing the metric – propagate at the speed of light as waves in free space.

4.4 Plane wave solutions for the metric perturbations

We now explore a little further the properties of solutions to equation (4.38). The simplest solutions are **plane waves**

$$\overline{h}_{\mu\nu} = \mathrm{Re}\left[A_{\mu\nu}\,\exp\left(ik_\alpha x^\alpha\right)\right] \tag{4.41}$$

where 'Re' denotes the real part, and the constant components $A_{\mu\nu}$ and k_α are known as the **wave amplitude** and **wave vector** respectively. (Note that, as it appears in equation (4.41), the k_α are the components of a one-form. However, since we are considering the weak field limit of a background Minkowski spacetime, converting between covariant and contravariant components is very straightforward).

Equation (4.41) may appear to restrict the metric perturbations to a particular mathematical form, but *any* $\overline{h}_{\mu\nu}$ can be Fourier-expanded as a superposition of plane waves.

The wave amplitude and wave vector components are not completely arbitrary. Firstly, $A_{\mu\nu}$ is symmetric, since $\overline{h}_{\mu\nu}$ is symmetric. This immediately reduces the number of indpendent components from 16 to 10.

Next, given that

$$\overline{h}_{\mu\nu,\alpha}{}^{,\alpha} = \eta^{\alpha\sigma}\,\overline{h}_{\mu\nu,\alpha\sigma} = 0 \tag{4.42}$$

it is easy to show that

$$k_\alpha\, k^\alpha = 0 \tag{4.43}$$

i.e. the wave vector is a **null vector**.

Thus, equation (4.41) describes a plane wave of frequency

$$\omega = k^t = \left(k_x^2 + k_y^2 + k_z^2\right)^{1/2} \tag{4.44}$$

propagating in direction $(1/k^t)\,(k_x, k_y, k_z)$.

Also, it follows from the Lorentz gauge condition

$$\overline{h}^{\mu\alpha}{}_{,\alpha} = 0 \tag{4.45}$$

that

$$\left(\overline{h}^{\alpha}_{\mu}\right)_{,\alpha} = 0 \tag{4.46}$$

from which condition it follows that

$$A_{\mu\alpha}\, k^\alpha = 0 \tag{4.47}$$

i.e. the wave amplitude components must be orthogonal to the wave vector, **k**.

Equation (4.47) is, in fact, four linear equations – one for each coordinate. This further reduces the number of independent components of $A_{\mu\nu}$ from 10 to 6. Can we restrict the components of the wave amplitude further still? The answer is 'yes', since we have sufficient freedom in our choice of gauge transformation.

It can be shown that, if we begin with arbitrary metric perturbation components, $h_{\mu\nu}^{(\text{old})}$ (defined on a background Minkowski spacetime), we can transform these components to

$$h_{\mu\nu}^{(\text{new})} \;=\; h_{\mu\nu}^{(\text{old})} \;-\; \xi_{\mu,\nu} \;-\; \xi_{\nu,\mu} \tag{4.48}$$

defined such that

$$G_{\mu\nu} \;=\; h_{\mu\nu,\alpha}^{(\text{new}),\alpha} \tag{4.49}$$

This transformation to the Lorentz gauge requires identification of 'vector' components ξ^{μ} which satisfy

$$\left(-\frac{\partial^2}{\partial t^2} \;+\; \nabla^2 \right) \xi^{\mu} \quad = \quad \overline{h}^{(\text{old})\,\mu\nu}{}_{,\nu} \tag{4.50}$$

Note that the transformation defined by equations (4.48) and (4.50) does **not** determine ξ^{μ} uniquely. To any set of components ξ^{μ} which satisfy equation (4.50), we could add the components, ζ^{μ}, to define a new transformation

$$\xi'^{\mu} = \xi^{\mu} + \zeta^{\mu} \tag{4.51}$$

and provided the ζ^{μ} satisfy

$$\left(-\frac{\partial^2}{\partial t^2} + \nabla^2 \right) \zeta^{\mu} = 0 \tag{4.52}$$

then ξ'^{μ} will still satisfy equation (4.50), and thus still express the Einstein tensor in the simplified, Lorentz gauge form of equation (4.49).

Equation (4.52) gives us four additional equations with which we can adjust the components of our gauge transformation, in order to choose a coordinate system which makes $h_{\mu\nu}$ – and hence $A_{\mu\nu}$ – as simple as possible. We can, in this way, reduce the number of independent components of $A_{\mu\nu}$ to only 2.

In fact, it can be shown (see e.g. Green Schutz, pg. 216) that the freedom we retain in our choice of ξ^{μ}, while still satisfying the Lorentz gauge conditions, allows us to restrict further $A_{\mu\nu}$ to satisfy

$$A^{\mu}_{\mu} = \eta^{\mu\nu} A_{\mu\nu} = 0 \tag{4.53}$$

and

$$A_{\alpha\beta}u^{\beta} = 0 \tag{4.54}$$

where u^{β} are the components of some constant unit four-vector. (The usefulness of introducing this restriction on $A_{\mu\nu}$ will become clear shortly). This choice of gauge transformation is known as the **Transverse – Traceless gauge**.

Suppose we now take ourselves to the background Lorentz frame in which u^{β} has components $(1, 0, 0, 0)$ – i.e.

$$u^{\beta} = \delta^{\beta}_{t} \tag{4.55}$$

Equations (4.54) and (4.55) then imply that

$$A_{\alpha t} = 0 \quad \text{for all } \alpha \tag{4.56}$$

Next we orient our spatial coordinate axes so that the wave is travelling in the positive z-direction, i.e.

$$k^{t} = \omega\,, \quad k^{x} = k^{y} = 0\,, \quad k^{z} = \omega \tag{4.57}$$

and

$$k_{t} = -\omega\,, \quad k_{x} = k_{y} = 0\,, \quad k_{z} = \omega \tag{4.58}$$

It then follows from equation (4.47) that

$$A_{\alpha z} = 0 \quad \text{for all } \alpha \tag{4.59}$$

In this coordinate frame, equation (4.41) simplifies to become

$$\overline{h}^{TT}_{\mu\nu} = A^{TT}_{\mu\nu} \cos\left[\omega(t-z)\right] \tag{4.60}$$

It is also straightforward to show, using equations (4.24) and (4.30), that $h_{\mu\nu}$ can be written in the form

$$h^{TT}_{\mu\nu} = B^{TT}_{\mu\nu} \cos\left[\omega(t-z)\right] \tag{4.61}$$

where the components $B^{TT}_{\mu\nu}$ are constant.

Equations (4.56) and (4.59), combined with the symmetry of $A_{\mu\nu}$, imply that the only non-zero components of $A_{\mu\nu}$ are A_{xx}, A_{yy} and $A_{xy} = A_{yx}$. Moreover, the traceless condition, equation (4.53), implies that $A_{xx} = -A_{yy}$. Hence, the components of $A_{\mu\nu}$ in the Transverse – Traceless gauge are

$$A^{TT}_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & A_{xx} & A_{xy} & 0 \\ 0 & A_{xy} & -A_{xx} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{4.62}$$

It then follows that

$$\overline{h}^{TT}_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \overline{h}^{TT}_{xx} & \overline{h}^{TT}_{xy} & 0 \\ 0 & \overline{h}^{TT}_{xy} & -\overline{h}^{TT}_{xx} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{4.63}$$

and

$$h^{TT}_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h^{TT}_{xx} & h^{TT}_{xy} & 0 \\ 0 & h^{TT}_{xy} & -h^{TT}_{xx} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{4.64}$$

4.5 The effect of gravitational waves on free particles

We see from equations (4.62) – (4.64) that the amplitude of the metric perturbation is described by just two independent constants, A_{xx} and A_{xy}. We can understand the physical significance of these constants by examining the effect of the gravitational wave on a free particle, initially in a wave-free region of spacetime.

Choose a background Lorentz frame in which the particle is initially at rest – i.e. the initial four-velocity of the particle is given by equation (4.55) – and choose the Transverse – Traceless Lorentz gauge so that the components $A^{TT}_{\mu\nu}$ in equation (4.41) are given by equation (4.62).

The particle's trajectory satisfies the geodesic equation

$$\frac{du^\beta}{d\tau} + \Gamma^\beta_{\mu\nu} u^\mu u^\nu = 0 \tag{4.65}$$

where τ is the proper time. Thus, the initial acceleration of the particle is

$$\left(\frac{du^\beta}{d\tau}\right)_0 = -\Gamma^\beta_{tt} = -\frac{1}{2}\eta^{\alpha\beta}\left(h_{\alpha t,t} + h_{t\alpha,t} - h_{tt,\alpha}\right) \tag{4.66}$$

However, from equation (4.56)

$$A_{\alpha t} = 0 \quad \Rightarrow \quad \overline{h}_{\alpha t} = 0 \tag{4.67}$$

Also, from equation (4.53)

$$A^\mu_\mu = 0 \quad \Rightarrow \quad \overline{h} = \overline{h}^\mu_\mu = 0 \tag{4.68}$$

Together, equations (4.67) and (4.68) imply that

$$h_{\alpha t} = 0 \quad \text{for all } \alpha \tag{4.69}$$

from which it follows immediately that

$$\left(\frac{du^\beta}{d\tau}\right)_0 = 0 \tag{4.70}$$

Hence a free particle, initially at rest, will remain at rest indefinitely. However, 'being at rest' in this context simply means that the *coordi-*

nates of the particle do *not* change. This is simply a consequence of our judicious choice of coordinate system, via the adoption of the Transverse – Traceless Lorentz gauge. As the gravitational wave passes, the coordinate system adjusts itself to the ripples in the spacetime, so that any particles remain 'attached' to their initial coordinate positions. Coordinates are merely frame-dependent labels, however, and do not directly convey any invariant geometrical information about the spacetime.

Suppose instead we consider the proper distance between two nearby particles, both initially at rest, in this coordinate system: one at the origin and the other at spatial coordinates $x = \epsilon$, $y = z = 0$. The proper distance between the particles is then given by

$$\Delta\ell = \int \left| g_{\alpha\beta} dx^{\alpha} dx^{\beta} \right|^{1/2} \tag{4.71}$$

i.e.

$$\Delta\ell = \int_0^{\epsilon} \left| g_{xx} \right|^{1/2} \quad \simeq \quad \sqrt{g_{xx}(x=0)}\,\epsilon \tag{4.72}$$

Now

$$g_{xx}(x=0) = \eta_{xx} + h_{xx}^{TT}(x=0) \tag{4.73}$$

so

$$\Delta\ell \simeq \left[1 + \frac{1}{2} h_{xx}^{TT}(x=0)\right] \epsilon \tag{4.74}$$

Since $h_{xx}^{TT}(x=0)$ is general is **not** constant, it follows that the proper distance between the particles will change as the gravitational wave passes. It is this change in the proper distance between 'test' particles which gravitational wave detectors attempt to measure.

4.5.1 The geodesic deviation of test particles

We can study the behaviour of test particles more formally using the idea of **geodesic deviation**, first introduced in GR-I. Let us define the vector ξ^α which connects the two particles introduced above. Then, for a weak gravitational field

$$\frac{\partial^2 \xi^\alpha}{\partial t^2} = R^\alpha_{\mu\nu\beta} u^\mu u^\nu \xi^\beta \tag{4.75}$$

where u^μ are the components of the four-velocity of the two particles. Since the particles are initially at rest, then

$$u^\mu = (1,0,0,0) \tag{4.76}$$

and

$$\xi^\beta = (0,\epsilon,0,0) \tag{4.77}$$

Equation (4.75) then simplifies to

$$\frac{\partial^2 \xi^\alpha}{\partial t^2} = \epsilon R^\alpha_{ttx} = -\epsilon R^\alpha_{txt} \tag{4.78}$$

Substituting from equation (4.24) for a weak gravitational field, we can write down the relevant components of the Riemann-Christoffel tensor in terms of the non-zero components of the metric perturbation

$$R^x_{txt} = \eta^{xx} R_{xtxt} = -\frac{1}{2} h^{TT}_{xx,tt} \tag{4.79}$$

$$R^y_{txt} = \eta^{yy} R_{ytxt} = -\frac{1}{2} h^{TT}_{xy,tt} \tag{4.80}$$

Hence, two particles initially separated by ϵ in the x-direction, have a geodesic deviation vector which obeys the differential equations

$$\frac{\partial^2}{\partial t^2} \xi^x = \frac{1}{2} \epsilon \frac{\partial^2}{\partial t^2} h^{TT}_{xx} \tag{4.81}$$

and

$$\frac{\partial^2}{\partial t^2} \xi^y = \frac{1}{2} \epsilon \frac{\partial^2}{\partial t^2} h^{TT}_{xy} \tag{4.82}$$

Similarly, it is straightforward to show that two particles initially separated by ϵ in the y-direction, have a geodesic deviation vector which obeys the differential equations

$$\frac{\partial^2}{\partial t^2} \xi^x = \frac{1}{2} \epsilon \frac{\partial^2}{\partial t^2} h^{TT}_{xy} \tag{4.83}$$

and

$$\frac{\partial^2}{\partial t^2}\,\xi^y \;=\; -\frac{1}{2}\epsilon\frac{\partial^2}{\partial t^2}h_{xx}^{TT} \tag{4.84}$$

4.5.2 Ring of test particles: polarisation of gravitational waves

We can further generalise equations (4.81) – (4.84) to consider the geodesic deviation of two particles – one at the origin and the other initially at coordinates $x = \epsilon\cos\theta$, $y = \epsilon\sin\theta$ and $z = 0$ – as a gravitational wave propagates in the z-direction. We can show that ξ^x and ξ^y obey the differential equations

$$\frac{\partial^2}{\partial t^2}\,\xi^x \;=\; \frac{1}{2}\epsilon\cos\theta\frac{\partial^2}{\partial t^2}h_{xx}^{TT} \;+\; \frac{1}{2}\epsilon\sin\theta\frac{\partial^2}{\partial t^2}h_{xy}^{TT} \tag{4.85}$$

and

$$\frac{\partial^2}{\partial t^2}\,\xi^y \;=\; \frac{1}{2}\epsilon\cos\theta\frac{\partial^2}{\partial t^2}h_{xy}^{TT} \;-\; \frac{1}{2}\epsilon\sin\theta\frac{\partial^2}{\partial t^2}h_{xx}^{TT} \tag{4.86}$$

Substituting from equations (4.61) and (4.64), we see that equations (4.85) and (4.86) have solution

$$\xi^x \;=\; \epsilon\cos\theta + \frac{1}{2}\epsilon\cos\theta\,B_{xx}^{TT}\cos\omega t \;+\; \frac{1}{2}\epsilon\sin\theta\,B_{xy}^{TT}\cos\omega t \tag{4.87}$$

and

$$\xi^y \;=\; \epsilon\sin\theta + \frac{1}{2}\epsilon\cos\theta\,B_{xy}^{TT}\cos\omega t \;-\; \frac{1}{2}\epsilon\sin\theta\,B_{xx}^{TT}\cos\omega t \tag{4.88}$$

Suppose we now vary θ between 0 and 2π, so that we are considering a circular ring of test particles, initially equidistant from the origin. Figure 12 (from Green Schutz) shows the effect of the passage of a plane gravitational wave on this ring of test particles. Panel (a) shows the particles before the wave reaches them.

Panel (b) shows the distortions produced by a wave for which $h_{xx}^{TT} \neq 0$ and $h_{xy}^{TT} = 0$. The upper and lower figures show two phases of the wave separated by π, at $t = t_1$ and $t = t_2$. When $t = t_1$, test particles on the x-axis are moved inwards; the gravitational wave reduces their proper distance from the origin. Particles on the y-axis, on the other hand, are moved outwards; the gravitational wave increases their distance from the origin. At time t_2 the behaviour is reversed as the oscillating component, h_{xx}^{TT}, has changed sign. Now particles on the x-axis are moving outwards in proper distance from the origin while particles on the y-axis are moving inwards.

Panel (c) shows the distortions produced by a wave for which $h_{xx}^{TT} = 0$ and $h_{xy}^{TT} \neq 0$. The upper and lower figures again show two phases of the wave separated by π, at $t = t_1$ and $t = t_2$. The behaviour of the ring of test particles in this case is the same as in panel (b), but simply rotated at 45° relative to that case.

Panels (b) and (c) of Figure 12 represent two distinct **polarisation states** for the gravitational wave; these states are rotated 45° with respect to one another. Contrast this with the two polarisation states of an electromagnetic wave, which are rotated by 90° relative to one another.

We can also see from Figure 12 that, at any instant, a gravitational wave is invariant under a rotation of 180° about its direction of propagation (in this case, the z-axis). By contrast, an electromagnetic wave is invariant under a rotation of 360°, and a *neutrino* wave is invariant under a rotation of 720°. We can understand this behaviour in terms of the *spin states* of the corresponding **gauge bosons**: the particles which are associated with the quantum mechanical versions of these waves.

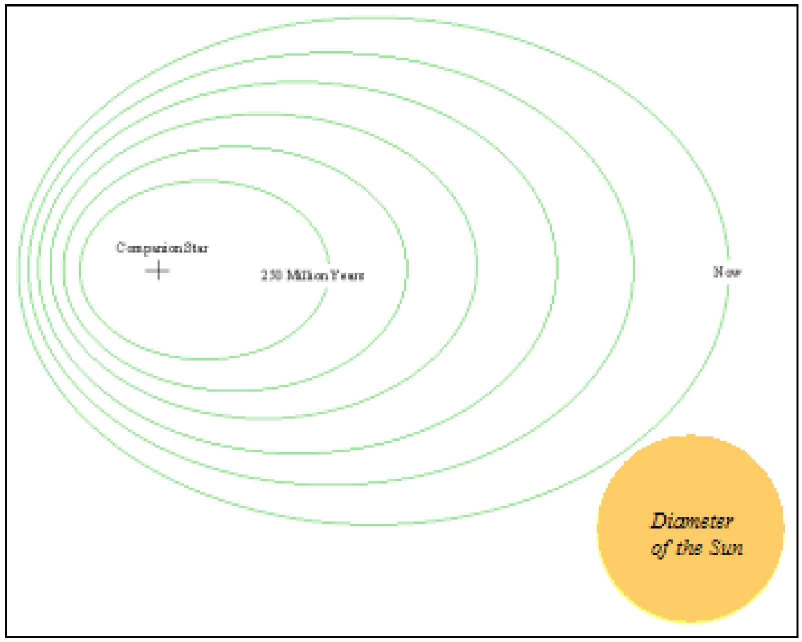
In general, the classical radiation field of a particle of spin, S, is invariant under a rotation of $360°/S$. Moreover, a radiation field of spin S

has precisely two independent polarisation states, which are inclined to each other at an angle of $90^\circ/S$. Thus, for an electromagnetic wave, corresponding to a photon of spin $S = 1$, the independent polarisation modes are inclined at 90° to each other.

We can, therefore, deduce from the inclination of the gravitational wave polarisation states, that the **graviton** (which is, as yet undiscovered, since we do not yet have a fully developed theory of quantum gravity!) must be a spin $S = 2$ particle. The fact that electromagnetic waves correspond to a spin $S = 1$ field and gravitational waves correspond to a spin $S = 2$ field is also intimately connected to their mathematical description in terms of geometrical objects: spin $S = 1$ fields are **vector fields**, which is why we require only a vector description for the electromagnetic field; spin $S = 2$ fields, on the other hand, are **tensor fields**, which is why we required to introduce tensors to describe the properties of the gravitational field.

Figure 11

Change in the semi-major axis of the orbit of the binary pulsar, PSR 1913+16, resulting from energy loss due to gravitational radiation.



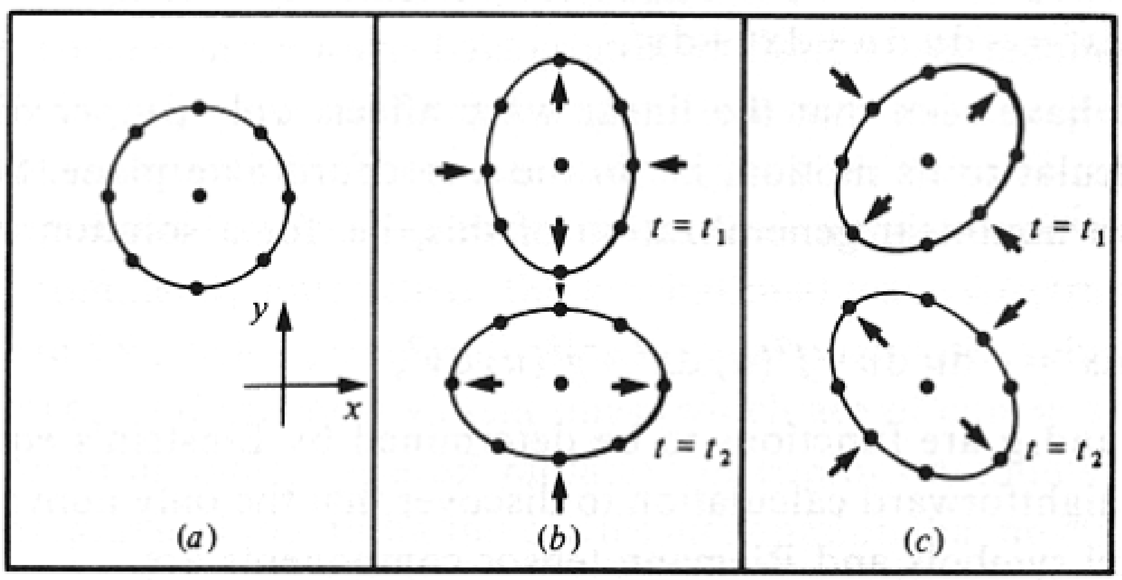


Figure 12

The two 'polarisation modes' of plane gravitational waves. See notes for an explanation of the different panels.

4.6 The amplitude of gravitational waves

How strong is gravitational radiation? In the previous sections we have developed a theoretical understanding of plane gravitational waves within the framework of a 'weak' gravitational field. This would tend to suggest that the effects of gravitational radiation should also be weak, but how can we quantify what we mean by 'weak' in this context?

A proper, quantitative treatment of this question lies well beyond the scope of this course, as it would require us to model in detail the astrophysical processes which produce gravitational waves. We will only address one particular aspect of this topic: in the next section we will show that gravitational waves are *quadrupolar* in nature.

Nevertheless, regardless of exactly *how* gravitational waves are produced, a very simplistic argument shows us that – even for 'strong' gravitational metric perturbations produced by by a distant source – one would expect the amplitude of gravitational waves incident at the Earth to be very small. Essentially, the argument requires us only to invoke a very general property of wave phenomena: that the amplitude of the wave is inversely proportional to the distance from the source.

A 'strong' metric perturbation would have $|h_{\alpha\beta}| \simeq 1$, in equation (4.1); we could expect perturbations of this amplitude only very close to the source, where the Newtonian potential (in geometrised units – see e.g. equation 2.89) is of order unity. For a source of mass, M (again, in geometrised units), this would occur at distances of order M from the source.

Hence, even if $|h_{\mu\nu}| \sim 1$ when $r \sim M$, it follows that, at distance, R, from a source of metric perturbations

$$|h_{\mu\nu}| \sim \frac{M}{R} \tag{4.89}$$

Consider, for example, the metric perturbations generated by the formation of a 8 solar mass black hole in e.g. the Andromeda galaxy – at a distance of $R \simeq 2 \times 10^{22}$ metres. The mass of the black hole, expressed in metres, is

$$M = 1.6 \times 10^{31}\,\mathrm{kg} \quad \simeq \quad 1.2 \times 10^{4}\,\mathrm{m} \tag{4.90}$$

Hence, even if the black hole produced strong metric perturbations, with $|h_{\mu\nu}| \sim 1$ at distances of order M, upon reaching the Earth these metric perturbations would have reduced in amplitude to

$$h_{\mu\nu} \sim 6 \times 10^{-19} \tag{4.91}$$

This number would also be of order the size of the fractional change in proper distance experienced by our ring of test particles as the metric perturbation passes. Moreover, this very simplistic calculation is a gross *over*-estimate, since it was based on assuming that the metric pertubations close to the source were of order unity. More realistic calculations – including detailed physical modelling (e.g. core collapse of a massive star, or coalescence of binary neutron stars) of the process which produces the metric perturbations – will lead to very much smaller amplitude perturbations close to the source, and hence correspondingly smaller perturbations at the Earth.

Thus, we see that the detection of gravitational waves from even relatively nearby (in cosmological terms) sources presents enormous technological challenges.

4.7 The quadrupolar nature of gravitational waves

We can understand something important about the nature of gravitational radiation by drawing analogies with the formulae that describe electromagnetic radiation. This approach only provides a crude, order-of-magnitude, estimate of the power of gravitational radaition, since the electromagnetic field is a vector field while the gravitational field is a tensor field, but it is good enough for our purposes in this course. Essentially, we will take familiar electromagnetic radiation formulae and simply replace the terms which involve the Coulomb force by their gravitational analogues from Newtonian theory.

4.7.1 Electric and magnetic dipoles

In electromagnetic theory, the dominant form of radiation from a moving charge or charges is **electric dipole radiation**. For a single particle (e.g. an electron) of charge, e, with acceleration, $\mathbf{a}$, and dipole moment changing as $\ddot{\mathbf{d}} = e\,\ddot{\mathbf{x}} = e\,\mathbf{a}$, the power output, or luminosity, is given by

$$L_{\text{electric dipole}} \propto e^2\,\mathbf{a}^2 \tag{4.92}$$

For a general distribution of charges, with net dipole moment, $\mathbf{d}$, the luminosity is

$$L_{\text{electric dipole}} \propto e^2\, \ddot{\mathbf{d}}^2 \tag{4.93}$$

The next strongest types of electromagnetic radiation are **magnetic dipole** and **electric quadrupole radiation**. For a general distribution of charges, the luminosity arising from magnetic dipole radiation is proportional to the second time derivative of the magnetic dipole moment, i.e.

$$L_{\text{magnetic dipole}} \propto \ddot{\mu} \tag{4.94}$$

where μ is given by a sum (or integral) over a distribution of charges:-

$$\mu \;=\; \sum_{q_i} \text{(position of } q_i) \times \text{(current due to } q_i) \tag{4.95}$$

4.7.2 Gravitational analogues

The gravitational analogue of the electric dipole moment is the **mass dipole moment**, $\mathbf{d}$, summed over a distribution of particles, $\{A_i\}$

$$\mathbf{d} \;=\; \sum_{A_i} m_i \mathbf{x}_i \tag{4.96}$$

where m_i is the rest mass and $\mathbf{x}_i$ is the position of particle A_i.

By analogy with equation (4.94), the luminosity of gravitational 'mass dipole' radiation should be proportional to the second time derivative of $\mathbf{d}$. However, the *first* time derivative of $\mathbf{d}$ is

$$\dot{\mathbf{d}} \;=\; \sum_{A_i} m_i \dot{\mathbf{x}}_i \quad \equiv \quad \mathbf{p} \tag{4.97}$$

where p is the **total linear momentum** of the system. Since the total momentum is conserved, it then must follow that the gravitational 'mass dipole' luminosity is zero – i.e **there can be no mass dipole radiation from any source**.

Similarly, the gravitational analogue of the magnetic dipole moment is

$$\mu \;=\; \sum_{A_i} (\mathbf{x}_i) \times (m_i \mathbf{v}_i) \quad \equiv \quad \mathbf{J} \tag{4.98}$$

where $\mathbf{J}$ is the **total angular momentum** of the system. Since the total angular momentum is *also* conserved, again it follows that the gravitational analogue of magnetic dipole radiation must have zero luminosity. Hence **there can be no dipole radiation of any sort from a gravitational source**.

Thus, the simplest form of gravitational radiation which has non-zero luminosity is **quadrupolar** in nature. We do not consider the mathematical details of quadrupolar radiation here, save to point out that it can be shown that the quadrupole from a **spherically symmetric mass distribution** is identically zero. This suggests an important result: that, at least up to quadrupole order, **metric perturbations which are spherically symmetric do not produce gravitational radiation**. Thus, if e.g. the collapse of a massive star is spherically symmetric, it will generate *no* gravitational waves.

In fact, it is possible to prove that this result is *also* true for higher order radiation (e.g. octupole etc.), although the proof is very technical and is not discussed further. Interested readers are referred to Chapters 9 and 10 of Green Schutz.

4.8 Example: a binary neutron star system

To close this chapter we consider the example of the gravitational wave signature of a particular (and rather extreme!) astrophysical system: a binary neutron star.

In general it can be shown (see, e.g. Green Schutz) that in the so-called *slow motion approximation* for a weak metric perturbation $h_{\mu\nu} << 1$

then for a source at distance r

$$h_{\mu\nu} = \frac{2G}{c^4 r} \ddot{I}_{\mu\nu} \tag{4.99}$$

where $I_{\mu\nu}$ is the **reduced quadrupole moment** defined as

$$I_{\mu\nu} = \int \rho(\vec{r}) \left(x_\mu x_\nu - \frac{1}{3} \delta_{\mu\nu} r^2 \right) dV \tag{4.100}$$

Consider a binary neutron star system consisting of two stars both of Schwarzschild mass M, in a circular orbit of coordinate radius R and orbital frequency f. For simplicity we define our coordinate system so that the orbital plane of the pulsars lies in the $x - y$ plane, and at coordinate time $t = 0$ the two pulsars lie along the $x-$axis. Substituting into equation (4.100)[1] it is then straightforward to show that

$$I_{xx} = 2MR^2 \left[\cos^2(2\pi f t) - \frac{1}{3} \right] \tag{4.101}$$

$$I_{yy} = 2MR^2 \left[\sin^2(2\pi f t) - \frac{1}{3} \right] \tag{4.102}$$

$$I_{xy} = I_{yx} = 2MR^2 \left[\cos(2\pi f t) \, \sin(2\pi f t) \right] \tag{4.103}$$

From equations (4.99) and (4.101) – (4.103) it then follows that

$$h_{xx} = -h_{yy} = h \cos\left(4\pi f t\right) \tag{4.104}$$

[1] taking the mass density distribution to be a sum of dirac delta functions – i.e. treating the pulsars as point masses

and

$$h_{xy} = h_{yx} = -h \sin(4\pi f t) \tag{4.105}$$

where the amplitude term h is given by

$$h = \frac{32\pi^2 G M R^2 f^2}{c^4 r} \tag{4.106}$$

We see from equation (4.106) that the binary system emits gravitational waves at twice the orbital frequency of the neutron stars. It is easy to verify that h in this equation is a dimensionless quantity; it is known as the **dimensionless strain**, and is related directly to the fractional change in arm length of e.g. a laser interferometric gravitational wave detector.

How large is h for a typical source? Suppose we take M equal to the Chandrasekhar mass, $M \sim 1.4 M_{\rm solar} = 2.78 \times 10^{30}$kg. We can then evaluate the constants in equation (4.106) and express h in more convenient units as

$$h = 2.3 \times 10^{-28} \, \frac{R^2[\text{km}] f^2[\text{Hz}]}{r[\text{Mpc}]} \tag{4.107}$$

If we take $R = 20$km, say, $f = 1000$Hz (which is approx. the frequency that Newtonian gravity would predict) and $r = 15$Mpc (corresponding to a binary system in e.g. the Virgo cluster), then we find that $h \sim 6\times10^{-21}$. Thus we see – just as we found in Section 5.6 from a more simplistic estimation – that the strain produced by a typical gravitational wave source places extreme demands upon detector technology.

Chapter 5

GR and Cosmology

In this final chapter we discuss some applications of general relativity to cosmology, providing a rigorous derivation of some results that are quoted and then discussed in detail within the final section of the course. We begin with a derivation of the general form of the spacetime metric which is consistent with the **cosmological principle**. We will then go on to use this metric to describe cosmological events.

5.1 The Robertson-Walker metric

In Part 1 the concept of the cosmological principle was introduced. Loosely speaking this states that, although the Universe is evolving, at an given time it appears **isotropic** and **homogeneous** on large scales.

In other words, provided we sample sample a large enough volume of the Universe to be truly representative, then things look pretty much the same everywhere and in all directions. The observational evidence in support of the cosmological principle comes chiefly from the cosmic microwave background radiation and galaxy redshift surveys, as discussed in the A1Y and honours cosmology courses.

Of course what we learn from the study of relativity is that the rigid Newtonian concepts of space and time are rather misleading and poorly defined, and so we need to consider how to formulate the cosmological principle within the framework of a fully covariant theory of general relativity. Essentially we want to translate the physical idea underpinning the cosmological principle – isotropy and homogeneity of the Universe – into a mathematical statement about the geometry of the Universe, as embodied by its metric. More specifically we want to derive the general form of the spacetime metric which is consistent with the cosmological principle. This metric is known as the **Robertson-Walker metric**.

5.1.1 Spherically symmetric spacetimes

Let the coordinates of our four-dimensional spacetime manifold have labels $\{x^0, x^1, x^2, x^3\}$. Then the interval between neighbouring events is given by

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \tag{5.1}$$

where, as usual, the summation convention runs over $0, 1, 2, 3$ for both indices μ and ν. In the language of GR, then, the cosmological principle means that we can slice up the four-dimensional geometry of spacetime by identifying a coordinate (let it be x^0 for reasons that are, or will soon become, obvious) such that hypersurfaces of constant x^0 have the following properties:

1. they are spherically symmetric (this is what we mean by isotropy within a GR framework)

2. they have the same value of the Ricci curvature scalar, R, at every point on the hypersurface (this is what we mean by homogeneity within a GR framework)

What considerations should influence our choice of the remaining hypersurface coordinates $\{x^1, x^2, x^3\}$? A sensible choice is **comoving coordinates**, which we first introduced in A1Y cosmology. In such a coordinate system, although the proper distance between observers changes as the background Universe expands, our comoving coordinates *also* expand in perfect step with the background. Thus, if we ignore any motions of galaxies *within* the hypersurfaces of constant x^0 (what we termed in Part 1 galaxy **peculiar motions**)[1], the comoving coordinates of each galaxy will remain fixed.

We now formally identify the x^0 coordinate as the proper time experienced by an observer in a galaxy – i.e. the time measured by a clock not moving relative to that observer, as would be the case for a comoving coordinate system.

Consider two galaxies with comoving coordinates $\{x^1, x^2, x^3\}$ and $\{x^1 + dx^1, x^2 + dx^2, x^3 + dx^3\}$. Suppose at some time t_0 the hypersurface of constant t_0 has line element

$$d\ell^2 = g_{ij}(t_0) dx^i dx^j \tag{5.2}$$

[1] In the interests of expediency we are being a little sloppy here, since strictly speaking we would consider neighbouring surfaces of *different* x^0 in order to define these galaxy motions

(i.e. this is just the interval from equation (5.1) with $dx^0 = 0$).

Suppose that at some later time, t_1 say, the hyperfsurface line element is given by

$$d\ell^2 = g_{ij}(t_1) dx^i dx^j \tag{5.3}$$

Since the galaxies have constant comoving coordinates, the change in the proper distance between the galaxies due to the expansion of the Universe between t_0 and t_1 must be described by the change in the metric coefficients g_{ij} between t_0 and t_1. In general the g_{ij}'s would depend not just on x^0 (i.e. t_0 or t_1) but would also be functions of the spatial coordinates x^1, x^2 and x^3. However, because of isotropy, the g_{ij}'s must all change at the same rate for $i = 1, 2, 3$ or else the expansion of the Universe would have a preferred direction. This means that the metric coefficients must depend only on x^0 (which we will henceforth label simply as t). Thus we can write the line element as

$$d\ell^2 = a^2(t) h_{ij} dx^i dx^j \tag{5.4}$$

where $a(t)$ is an overall scale factor (known generally as the **cosmic scale factor**) that depends only on time t, and h_{ij} are the metric coefficients for the comoving coordinates on the three-dimensional spatial hypersurface at constant t. Note that the h_{ij} do *not* depend on t, since they describe the metric coefficients for comoving coordinates.

The full spacetime interval of equation (5.1) can therefore be written as

$$ds^2 \quad = \quad -dt^2 + g_{0i}dtdx^i + a^2(t)h_{ij}dx^i dx^j \tag{5.5}$$

We can use symmetry arguments, similar to those of Section 2.3, to show that

$$g_{0i} = \vec{e}_0 \cdot \vec{e}_i = 0 \tag{5.6}$$

so that

$$ds^2 \quad = \quad -dt^2 + a^2(t)h_{ij}dx^i dx^j \tag{5.7}$$

We know that $g_{00} = -1$ because we have identified t as the proper time, τ, along a line with $dx^i = 0$ and

$$d\tau = \sqrt{-ds^2} \tag{5.8}$$

What form should the h_{ij} take? Since they are independent of t, and should describe a spatial hypersurface which is spherically symmetric, we can make use of the results which we already derived in Chapter 1. There we showed that the metric of a static, spherically symmetric spacetime took the form of equation (1.23)

$$ds^2 = -e^\nu dt^2 + e^\lambda dr^2 + r^2 \left(d\theta^2 + \sin^2\theta \, d\phi^2\right) \tag{5.9}$$

where $\lambda \equiv \lambda(r)$. In the present case, then, the metric of our spherically symmetric spacetime takes the form

$$ds^2 \quad = \quad -dt^2 + a^2(t)\left[e^\lambda dr^2 + r^2\left(d\theta^2 + \sin^2\theta \, d\phi^2\right)\right] \tag{5.10}$$

(Note that the g_{00} component appears in many textbooks as $-c^2$, but recall that we are generally working with geometrised units in which $c = 1$).

We can determine the function $\lambda(r)$ by using our requirement that spatial hypersurfaces of constant t have constant curvature. For the line element

$$d\ell^2 = e^\lambda dr^2 + r^2\left(d\theta^2 + \sin^2\theta \, d\phi^2\right) \tag{5.11}$$

(i.e. for simplicity taking $a(t) = 1$) we can write down an expression for the curvature scalar, R, using results which we derived in Chapter 1. Firstly the Christoffel symbols are the same as those derived in equations (1.24), but with $\nu' = \nu'' = 0$. Thus

$$
\begin{array}{lclclcl}
\Gamma^r_{rr} & = & \frac{1}{2}\lambda' & & \Gamma^\theta_{r\theta} = \Gamma^\theta_{\theta r} & = & \frac{1}{r} \\
\Gamma^r_{\theta\theta} & = & -re^{-\lambda} & & \Gamma^\theta_{\phi\phi} & = & -\sin\theta\cos\theta \\
\Gamma^r_{\phi\phi} & = & -re^{-\lambda}\sin^2\theta & & \Gamma^\phi_{r\phi} = \Gamma^\phi_{\phi r} & = & \frac{1}{r} \\
\Gamma^\phi_{\theta\phi} = \Gamma^\phi_{\phi\theta} & = & \cot\theta & & & &
\end{array}
$$

$$
\text{All others zero} \tag{5.12}
$$

Similarly, adapting the results of equations (1.27) – (1.29), the only non-zero components of the Ricci tensor are

$$
R_{rr} = -\frac{\lambda'}{r} \tag{5.13}
$$

$$
R_{\theta\theta} = 1 - e^{-\lambda}(1 - \lambda') \tag{5.14}
$$

$$
R_{\phi\phi} = R_{\theta\theta}\sin^2\theta \tag{5.15}
$$

Since both the metric and Ricci tensors are orthogonal, the curvature scalar is given by

$$
R = g^{ij}R_{ij} = g^{rr}R_{rr} + g^{\theta\theta}R_{\theta\theta} + g^{\phi\phi}R_{\phi\phi} \tag{5.16}
$$

and we then find that

$$
R = 2\left[\frac{1}{r^2} + \frac{\lambda' e^{-\lambda}}{r} - \frac{e^{-\lambda}}{r^2}\right] \tag{5.17}
$$

or equivalently

$$R = \frac{2}{r^2}\left[1 - e^{-\lambda}\left(1 - r\lambda'\right)\right] \tag{5.18}$$

A little further algebra shows that equation (5.18) can be re-written as

$$R = \frac{2}{r^2}\left[1 - \frac{d}{dr}\left(re^{-\lambda}\right)\right] \tag{5.19}$$

Requiring that the curvature is equal to the same constant everywhere on the spatial hypersurface implies that

$$\frac{d}{dr}\left(re^{-\lambda}\right) = 1 - \frac{1}{2}Cr^2 \tag{5.20}$$

where C is a constant (equal to the value of the curvature scalar). Integrating equation (5.20)

$$re^{-\lambda} \quad = \quad r - \frac{1}{6}Cr^3 + A \tag{5.21}$$

or

$$e^{-\lambda} \quad = \quad 1 - \frac{1}{6}Cr^2 + \frac{A}{r} \tag{5.22}$$

where A is another constant.

Since spacetime is locally flat we require that $e^{\lambda} \rightarrow 1$ as $r \rightarrow 0$. Hence it also follows that $e^{-\lambda} \rightarrow 1$ as $r \rightarrow 0$. This implies that $A = 0$. Re-writing

$C/6$ as a new constant k, finally gives us

$$e^{\lambda} = \frac{1}{1 - kr^2} \tag{5.23}$$

which means that the metric of the full four-dimensional spacetime may be written as

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1 - kr^2} + r^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right)\right] \tag{5.24}$$

This is the most general form of the spacetime metric which has the property that its spatial hypersurfaces are isotropic and homogenous – in concordance with the cosmological principle. The metric is known as the **Robertson-Walker metric**. The values $k = 0$, $k > 0$ and $k < 0$ correspond to spatial hypersurfaces which have zero, positive and negative curvature respectively. This should not be surprising, since k is proportional to the constant, C, equal to the value of the curvature scalar, R, everywhere on the spatial hypersurface.

In fact we need only ever consider three values of k: $k = 0$, $k = +1$ and $k = -1$. This is because the line element for any other positive (negative) value of k can be reduced to the line element for $k = +1$ ($k = -1$) by a simple re-scaling of the cosmic scale factor, $a(t)$. The values $k = 0$, $k = +1$, and $k = -1$ are said to represent Universes with flat, closed and open geometries respectively.

5.2 Friedmann-Lemaitre Models

Our next task will be to explore cosmologically interesting model Universes which are described by the Robertson-Walker metric. These are known as Friedmann–Lemaitre models, after the researchers who first studied them. The steps required to derive these models are as follows

- Compute the Einstein tensor for the Robertson-Walker metric (by first computing its Christoffel symbols and hence its Riemann-Christoffel tensor)

- Assume a form for the energy-momentum tensor which is consistent with the cosmological principle

- Apply Einstein's equations to connect the Einstein and energy-momentum tensors

- Solve Einstein's equations to derive expressions for the time evolution of the cosmic scale factor and other important physical quantities

5.2.1 Christoffel symbols for the Robertson-Walker metric

Since the Robertson-Walker metric is orthogonal, the contravariant metric components are just the inverse of the covariant metric components. Thus

$$g^{tt} = -1; \quad g^{rr} = \frac{1 - kr^2}{a^2(t)}; \quad g^{\theta\theta} = \frac{1}{a^2(t)r^2}; \quad g^{\phi\phi} = \frac{1}{a^2(t)r^2 \sin^2\theta} \tag{5.25}$$

The orthogonality of the metric means that we can also use the results of equations (1.12) to simplify the calculation of its Christoffel symbols. Indeed equations (5.12) already give us the Christoffel symbols for all terms not involving t, although those expressions were given in terms of the function $\lambda(r)$. Note, however, that

$$\lambda = \ln\left(\frac{1}{1 - kr^2}\right) = -\ln\left(1 - kr^2\right) \tag{5.26}$$

so that

$$\lambda' = \frac{2kr}{1 - kr^2} \tag{5.27}$$

It is then straightforward, although *very* laborious, to show that the Christoffel symbols for the Robertson-Walker metric are

$$
\begin{array}{lclclcl}
\Gamma^t_{rr} & = & \frac{a\dot{a}}{1-kr^2} & & \Gamma^t_{\theta\theta} & = & a\dot{a}r^2 \\
\Gamma^t_{\phi\phi} & = & a\dot{a}r^2\sin^2\theta & & \Gamma^r_{rr} & = & \frac{kr}{1-kr^2} \\
\Gamma^r_{rt} = \Gamma^r_{tr} & = & \frac{\dot{a}}{a} & & \Gamma^r_{\theta\theta} & = & -r\left(1-kr^2\right) \\
\Gamma^r_{\phi\phi} & = & -r\left(1-kr^2\right)\sin^2\theta & & \Gamma^\theta_{\phi\phi} & = & -\sin\theta\cos\theta \\
\Gamma^\phi_{\theta\phi} = \Gamma^\phi_{\phi\theta} & = & \cot\theta & & \Gamma^\theta_{r\theta} = \Gamma^\theta_{\theta r} & = & \frac{1}{r} \\
\Gamma^\phi_{r\phi} = \Gamma^\phi_{\phi r} & = & \frac{1}{r} & & \Gamma^\theta_{\theta t} = \Gamma^\theta_{t\theta} & = & \frac{\dot{a}}{a} \\
\Gamma^\phi_{\phi t} = \Gamma^\phi_{t\phi} & = & \frac{\dot{a}}{a} & & & &
\end{array}
$$

$$
\text{All others zero} \tag{5.28}
$$

5.2.2 The Ricci tensor for the Robertson-Walker metric

Using the fact that the Ricci tensor is a contraction of the Riemann-Christoffel tensor, we see that e.g.

$$
R_{tt} = R^\sigma_{t\sigma t} = R^t_{ttt} + R^r_{trt} + R^\theta_{t\theta t} + R^\phi_{t\phi t} \tag{5.29}
$$

The first term on the right hand side is given by

$$
R^t_{ttt} = \Gamma^\sigma_{tt}\Gamma^t_{\sigma t} - \Gamma^\sigma_{tt}\Gamma^t_{\sigma t} + \Gamma^t_{tt,t} - \Gamma^t_{tt,t} \tag{5.30}
$$

each term of which, from equations (5.28), can be seen to be zero. After evaluation of the remaining terms on the right hand side of equation

(5.29) it is straightforward to show that

$$R_{tt} = -3\,\frac{\ddot{a}}{a} \tag{5.31}$$

Similar calculations then give

$$R_{rr} \quad = \quad \frac{2\dot{a}^2 + 2k + a\ddot{a}}{1 - kr^2} \tag{5.32}$$

$$R_{\theta\theta} \quad = \quad r^2\left[a\ddot{a} + 2\dot{a}^2 + 2k\right] \tag{5.33}$$

$$R_{\phi\phi} \quad = \quad R_{\theta\theta}\,\sin^2\theta \tag{5.34}$$

5.2.3 The Einstein tensor for the Robertson-Walker metric

Combining equation (5.25) and equations (5.31) – (5.34), and after some algebraic tidying up, it follows that the curvature scalar, $R = R_{\mu\nu}g^{\mu\nu}$, is given by

$$R \quad = \quad 6\left[\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} + \frac{k}{a^2}\right] \tag{5.35}$$

We can now evaluate the components of the Einstein tensor, in its fully covariant form, namely

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R \tag{5.36}$$

Since both the Ricci and metric tensors are orthogonal, so too is the

Einstein tensor and its non-zero components are found to be

$$G_{tt} = \frac{3\dot{a}^2}{a^2} + \frac{3k}{a^2} \tag{5.37}$$

$$G_{rr} = -\frac{a^2}{1-kr^2}\left[\frac{2\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} + \frac{k}{a^2}\right] \tag{5.38}$$

$$G_{\theta\theta} = -r^2a^2\left[\frac{2\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} + \frac{k}{a^2}\right] \tag{5.39}$$

$$G_{\phi\phi} = G_{\theta\theta}\sin^2\theta \tag{5.40}$$

5.2.4 The energy-momentum tensor for the cosmic fluid

To proceed further we need to assume an appropriate form for the energy-momentum tensor. It is customary to treat the cosmological matter and energy fields as a **perfect fluid**. From Part 2 the general expression for the components of the energy-momentum tensor, in fully covariant form, is

$$T_{\mu\nu} = (\rho + P)u_\mu u_\nu + Pg_{\mu\nu} \tag{5.41}$$

with $c = 1$. Here ρ and P denote the density and pressure of the fluid, and u_μ denotes the (covariant components of the) four-velocity of a fluid element. In order to be consistent with the cosmological principle, which requires that the Universe be homogeneous, ρ and P must be functions only of t.

For the case of a perfect fluid the energy-momentum tensor simplifies to

$$\mathbf{T} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & P & 0 & 0 \\ 0 & 0 & P & 0 \\ 0 & 0 & 0 & P \end{pmatrix} \tag{5.42}$$

Note that the perfect fluid approximation assumes that fluid elements do not exchange energy with their neighbours due to heat conduction or viscosity. (Such processes would produce the *shearing* of fluid elements, which would introduce non-zero off-diagonal terms in the energy-momentum tensor). This approximation breaks down when one considers e.g. the dynamics of hot gas in individual or interacting galaxies, but it is a remarkably good approximation when one considers the behaviour of the 'background' cosmic fluid on large scales. Indeed, since we believe that the matter content of the Universe is dominated today by weakly interacting cold dark matter, we can also usually neglect the contribution of matter pressure and further approximate the cosmic fluid as *dust*. In recent years, however, it has become fashionable to reintroduce in a more general way the notion of cosmic pressure via the so-called **equation of state parameter**, w, in the equation

$$P = w\rho \tag{5.43}$$

Different values of w then correspond to different physical regimes with correspondingly different sources of pressure: e.g. dust is the case $w = 0$; a radiation dominated Universe (which one can think of as an ideal 'photon' gas, with $c = 1$) corresponds to $w = 1/3$. More exotic cosmological models, such as networks of cosmic strings, can also be incorporated by appropriate choice of w, and a wide range of models – with evocative names such as **dark energy**, **quintessence** and even **phantom energy** – in which w evolves with time has also been proposed. We return briefly later to the idea of the w parameter, and its relation to the so-called **cosmological constant**. Although this exciting topic lies mainly well beyond the scope of this course, it is a highly active area of current cosmology research and we return to it a little in Part 4.

5.2.5 Solutions of Einstein's equations

Having armed ourselves with the components of the Einstein tensor and the energy-momentum tensor, we can now apply the solution of Einstein's equations, i.e.

$$8\pi G T_{\mu\nu} = G_{\mu\nu} \tag{5.44}$$

Note that the G on the left hand side is the Newtonian gravitational constant (which we now include explicitly, as is common practice in most cosmology texts – even those which take $c = 1$, as we do here) and the factor of $8\pi G$ arises from requiring that the solution to Einstein's equations reduces to the Newtonian expression in the limit of a weak gravitational field.

If we first consider the 'tt' components of the Einstein and energy-momentum tensors, it follows from equations (5.37) and (5.44) that

$$8\pi G\rho = \frac{3\dot{a}^2}{a^2} + \frac{3k}{a^2} \tag{5.45}$$

Re-arranging, this can be written as

$$H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} \tag{5.46}$$

where we have used the definition of the **Hubble parameter**, $H = \dot{a}/a$.

Equation (5.46) is identical to the so-called **Friedmann equation** derived in some elementary astronomy courses from purely Newtonian considerations, but now re-derived rigorously within the framework of GR. This is essentially a differential equation for the time-evolution of the cosmic scale factor; recall that we can use the Friedmann equation to identify a **critical density** for the Universe that marks the division between models which would expand indefinitely and models which would expand and then contract to a 'Big Crunch'.

Of course within our rigorous GR framework we can make use of the other components of the Einstein and energy-momentum tensors to derive a further constraint equation for the cosmic scale factor; this is usually referred to as the second Friedmann equation and may be presented in different forms in different textbooks – depending on which component(s) of the tensors are used to derive it. Here we obtain our second equation by *contracting* the Einstein and energy-momentum tensors – i.e. by computing

$$\mathcal{T} = g^{\mu\nu} T_{\mu\nu} \qquad \text{and} \qquad \mathcal{G} = g^{\mu\nu} G_{\mu\nu} \tag{5.47}$$

(Here we have introduced the notation $\mathcal{T}$ and $\mathcal{G}$ to avoid confusion with e.g. temperature and with the Newtonian gravitational constant respectively).

Once again, the orthogonality of the Einstein, energy-momentum and metric tensors considerably simplifies our calculations. It is easy to show that

$$\mathcal{G} = -6\,\frac{\ddot{a}}{a} - 6\,\frac{\dot{a}^2}{a^2} - 6\,\frac{k}{a^2} \tag{5.48}$$

and

$$\mathcal{T} = 3P - \rho \tag{5.49}$$

From Einstein's equations, then

$$8\pi G\mathcal{T} = \mathcal{G} \tag{5.50}$$

or

$$8\pi G(3P - \rho) = -6\,\frac{\ddot{a}}{a} - 6\,\frac{\dot{a}^2}{a^2} - 6\,\frac{k}{a^2} \tag{5.51}$$

Adding together twice equation (5.45) and equation (5.51) eliminates the terms involving $\dot{a}$ and k to give

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\,(\rho + 3P) \tag{5.52}$$

5.2.6 Einstein's cosmological constant

We can see from equation (5.52) that, for 'normal' matter with $\rho > 0$ and $P \geq 0$, it follows that we cannot have a static solution for $a(t)$, with $\dot{a} = \ddot{a} = 0$. It was essentially this fact which motivated Einstein to introduce his so-called *cosmological constant* since, prior to Hubble's discovery of the recession of the nebulae, the expectation was that the Universe was indeed static. Einstein introduced the cosmological constant, Λ, as an additional term in the Einstein tensor, i.e.

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + g_{\mu\nu} \Lambda \qquad (5.53)$$

or

$$G^{\mu\nu} = R^{\mu\nu} - \frac{1}{2} g^{\mu\nu} R + g^{\mu\nu} \Lambda \qquad (5.54)$$

Since the covariant derivatives of the metric are zero, it still follows that

$$G^{\mu\nu}{}_{;\nu} = 0 \qquad (5.55)$$

as before.

We can think of Λ as a modification to the large-scale geometry of the Universe, although nowadays it is generally more common to think of the cosmological constant instead as a modification of the energy-momentum tensor, rather than the Einstein tensor. In some sense both descriptions are equivalent, but the latter interpretation is perhaps more useful in helping to understand physically what the cosmological constant actually *is* – e.g. due to the zero-point energy of the vacuum.

Combining equation (5.54) with the earlier results of sections 7.2.3 and 7.2.4, it is straightforward to show that the Friedmann equations (5.46) and (5.52) now take the form

$$H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3} \tag{5.56}$$

and

$$\frac{\ddot{a}}{a} = -\frac{4}{3}\pi G\,(\rho + 3P) + \frac{\Lambda}{3} \tag{5.57}$$

The presence of the Λ terms means that a solution for $a(t)$ *can* be found with $\dot{a} = \ddot{a} = 0$. For example, if the Universe is assumed to be matter dominated with $P = 0$ (i.e. assuming a dust solution), then setting $\ddot{a} = 0$ in equation (5.57) implies that $\Lambda > 0$, since $\rho > 0$ for normal matter. This in turn implies that $k > 0$ if we set $H = 0$ in equation (5.56). Hence $k = 1$, and from equation (5.57)

$$\Lambda = 4\pi G\rho \tag{5.58}$$

and

$$a = \frac{1}{\sqrt{4\pi G\rho}} = \frac{1}{\sqrt{\Lambda}} \tag{5.59}$$

Equation (5.59) is known as the *Einstein solution* for the scale factor, and it does indeed satisfy what was in 1916 the apparently sensible requirement that the scale factor be constant, corresponding to a static Universe. However, it is possible to show that this solution is inherently *unstable* – i.e. the slightest perturbation of the density would lead to rapid expansion or collapse. The Einstein solution is rather like the condition of a pencil balanced on its point: although the pencil is static, the slightest disturbance would cause it to topple over. This weakness of the Einstein solution – together with the even stronger physical evidence provided by Hubble's observations of the expansion of the Universe – prompted Einstein much later to describe his introduction of the cosmological constant as his 'greatest blunder'. Indeed one might argue that it is unfortunate that Einstein did not stick to the courage of his convictions and use the result of equation (5.52) to make the startling prediction that the Universe *was* expanding. The confirmation of such a prediction by Hubble would have been an even more spectacular vindication of GR than, say, the observations of light deflection carried out by Eddington during a total Solar eclipse.